The dynamics and statistics of knots in biopolymers

Saeed Najafi

DISSERTATION

zur Erlangung des Grades
"Doktor der Naturwissenschaften"

Johannes Gutenberg-Universität Mainz
Fachbereich Physik, Mathematik, Informatik

September 2017

Angefertigt am:

Max-Planck-Institut für Polymerforschung

Abstract

Knots have a plethora of applications in our daily life from fishing to securing surgical sutures. Even within the microscopic scale, various polymeric systems have a great capability to become entangled and knotted. As a notable instance, knots and links can either appear spontaneously or by the aid of chaperones during entanglement in biopolymers such as DNA and proteins. Despite the fact that knotted proteins are rare, they feature different type of topologies. They span from simple trefoil knot, up to the most complex protein knot, the Stevedore. Knotting ability of polypeptide chain complicates the conundrum of protein folding, that already was a difficult problem by itself. The sequence of amino acids is the most remarkable feature of the polypeptide chains, that establishes a set of interactions and govern the protein to fold into the knotted native state. We tackle the puzzle of knotted protein folding by introducing a structure based coarse-grained model. We show that the nontrivial structure of the knotted protein can be encoded as a set of specific local interaction along the polypeptide chain that maximizes the folding probability. In contrast to proteins, knots in sufficiently long DNA and RNA filaments are frequent and diverse with a much smaller degree of sequence dependency. The presence of topological constraints in DNA and RNA strands give rise to a rich variety of structural and dynamical features. We show that the knotting probability of a dsDNA, can be increased by introducing, along its sequence, two adhesive regions. We show that entanglement pattern in links and knots, play a key role in conformational properties of chains. In particular, we demonstrate that a double knotted semi-flexible polymer chain under strong stretching possesses a free energy minimum when the two knots are intertwined, and that the free energy of the intertwined state is deepening when the relative chirality of the knots is opposite. Additionally, we show that the braid of DNA rings with identical/non-identical crossing pattern, enforces negative/positive and weak/strong correlation between the entangled rings.

Zusammenfassung

Knoten haben von der Fischerei bis zur Schließung chirurgischer Nähte eine Fülle von Anwendungen in unserem täglichen Leben. Auch auf mikroskopischer Skala haben verschiedene polymere Systeme die Fähigkeit sich zu verschränken und zu verknoten. Bemerkenswerterweise können Knoten und Verflechtungen entweder spontan oder durch die Hilfe von Chaperonen während der faltung in Biopolymeren wie DNA und Proteinen auftreten. Obwohl Knoten in Proteinen selten sind, haben Sie verschiedene Arten von Topologien. Diese reichen von einfachen Kleeblattknoten bis zu dem komplexesten Protein Knoten, dem Stevedore-Knoten. Die Fähigkeit der Polypeptide sich zu verknoten, erschwert das bereits komplexe Problem der Proteinfaltung zusätzlich. Die Aminoäuresequenz der Polypeptide ist das bedeutendste Merkmal, deren Wechselwirkungen das Falten in die verknoteten Grundzustände des Proteins leiten. Wir gehen das Problem der Knotenbildung durch die Einführung eines strukturbasierenden und grobkörnigen Modells an. Wir zeigen, dass eine nicht triviale Struktur des verknotenden Proteins als eine Reihe spezifischer lokaler Wechselwirkung entlang der Kette kodiert werden kann, um die Faltung zu maximieren. Im Gegensatz zu Proteinen sind Knoten in hinreichend langen DNA und RNA-Filamenten häufiger, vielfältiger, und mit einem viel geringeren Grad von der Sequenz abhängig. Das Vorhandensein von topologischen Randbedingungen in DNA und RNA-Stränge führt zu einer Fülle von strukturellen und dynamischen Merkmalen. Wir zeigen, dass die Wahrscheinlichkeit zur Knotenbildung von dsDNA durch die Einführung von zwei haftenden Regionen erhöht werden kann. Zudem zeigen wir, dass Kreuzungsmuster in Verflechtungen und Knoten eine Schlüsselrolle bei Konformationseigenschaften der Kette spielen. Insbesondere zeigen wir, dass eine doppelt verknotete semi-flexiblen Polymerkette unter starker Streckung sich in einem freien Energieminimum befindet, wenn die beiden Knoten miteinander verflochten sind, und dass die freie Energie des verflochtenen Zustandes niedriger ist, wenn die relative Chiralität der Knoten

gegensätzlich ist. Darüber hinaus zeigen wir, dass das Geflecht von DNA-Ringen mit identischen/nicht identischen Kreuzungsmustern eine negative/positive und schwach/starke Korrelation zwischen den verstrickten Ringen erzwingt.

Table of contents

Included Publications

The contents of this Thesis represents my doctoral studies at the Max Planck Institute for Polymer Research and has been submitted to Johannes Gutenberg University, Mainz as a cumulative dissertation. Chapter 1 is an introduction to polymer physics and topology in polymers. Results are presented in Chapters 2–6, which consist of the following published papers and manuscripts:

- **Chapter 2**

  Saeed Najafi and Raffaello Potestio "**Folding of small knotted proteins: Insights from a mean field coarse-grained model**" The Journal of Chemical Physics (2015) 143(24):243121

- **Chapter 3**

  Saeed Najafi and Raffaello Potestio "**Two Adhesive Sites Can Enhance the Knotting Probability of DNA**" PLoS ONE (2015) 10(7):e0132132

- **Chapter 4**

  Saeed Najafi, Luca Tubiana, Rudolf Podgornik, and Raffaello Potestio "**Chirality modifies the interaction between knots**" EPL (Europhysics Letters) (2016) 114(5)

- **Chapter 5**

  Saeed Najafi, Rudolf Podgornik, Raffaello Potestio, and Luca Tubiana "**Role of Bending Energy and Knot Chirality in Knot Distribution and Their Effective Interaction along Stretched Semiflexible Polymers**" Polymers (2016) 8(10):347

- **Chapter 6**

  Saeed Najafi and Raffaello Potestio "**Entanglement of knotted DNA ring and an entwined DNA loop**" In preparation

List of figures

List of tables

Chapter 1

Introduction to polymer physics and topology in polymers

1.1 Introduction

During polymerization, polymers with tremendous diversity in their structure and functionality form from covalently-bonded discrete subunits, generically called "*monomers*". The degree of polymerization that by definition is the number of monomers of the chain can determine the physical and chemical properties of the polymers [5]. Mechanisms that control structure, dynamics, and rheology of polymers are found to directly be affected by the type of monomers [1, 3, 5]. Liquid, glasses, crystals, and liquid crystals are some example of materials that polymers can form. In Fig. 1.1 panel a and c, we show polymethylmethacrylate and polystyrene that are some example of glassy polymers. Plastics and colors are some example of artificial polymers. Biological macromolecules such as proteins, DNA and RNA filaments are some example of bio-polymers [4, 5].

Polymeric materials with large variety of physical and chemical properties, have fascinated scientists for more than a century. There are many types of models that have been proposed to study the physics of polymers, where each model may correspond to specific physiological conditions [1, 2]. In this thesis we study linear chains with identical monomers. Although linear models of polymers may be an oversimplified version of complex heteropolymers such as DNA and proteins, they can still accurately describe the generic large-scale behavior of the system under

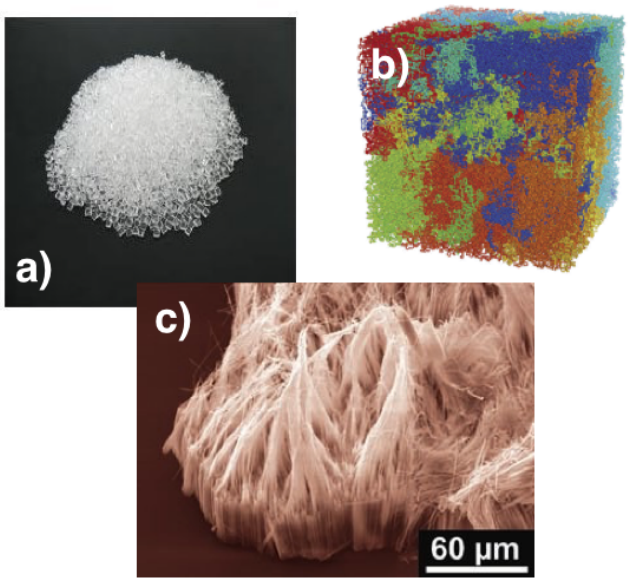


Fig. 1.1 In panel a and b, we show Poly(methyl methacrylate) (PMMA) that is a transparent thermoplastic glass polymer and a schematic view of globular polymer, respectively. In panel c we show electron microscopy image showing the segmented nature of polystyrene hybrid nano-fibers.

certain ambient conditions [6–9]. Here, we outline basic concepts of polymer physics and introduce the methods that one can use to probe the physical features of polymers.

1.2 Ideal and real chains

1.2.1 Ideal chain

The ideal chain model serves as an excellent foundation for polymer physics [1, 2]. In this model, there is no interaction between non-consecutive monomers of polymer chains. This simplification never completely characterized for real chains, but there are some situations in which the actual polymeric systems behavior resembles ideal chains. While this model is unrealistic, practically, due to the screening effect of the solvents with high concentrations of ions, nonbonded interactions between monomers are negligible [1].

In this section we introduce some models of ideal chains. In every model due to the various chemical structures of the polymer, different assumptions for bend and torsion angles are applied. However, in all systems modeled as an ideal chain, the interactions between monomers separated by large distance is ignored.

Conformations of ideal polymer chain

Consider a linear polymer that consists of N monomers. The vector $\vec{r}_i$ from $(i-1)th$ to ith monomer called a bond vector, and the sum over all bond vectors on polymer chain is known as the end-to-end distance of the polymer chain ($\vec{R}$):

$$\vec{R} = \sum_{i=1}^{N-1} \vec{r}_i \tag{1.1}$$

The mean square value of the end-to-end distance is given by:

$$< \vec{R}^2 >= \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} < \vec{r}_i.\vec{r}_j > \tag{1.2}$$

We assume that the distance between consecutive beads is constant and equal to a. Thus, the scalar product of the bond vectors can be written as $\vec{r}_i.\vec{r}_j = a^2 cos(\theta_{ij})$, where θ_{ij} is the angle between i and jth bond vectors as shown in Fig 1.3. Since there is no orientational correlation between the bonds in the free jointed chain, then the off-diagonal values of $< cos(\theta_{ij}) >$ are zero. However the diagonal values are 1, giving rise to $< R^2 >= (N-1)a^2$. As already mentioned, there is no orientational correlation in the ideal chain, meaning that one can consider it as a free jointed chain that consists of $N-1$ bonds with length of a, which its maximum end-to-end distance is $R_{max} = (N-1)a$.

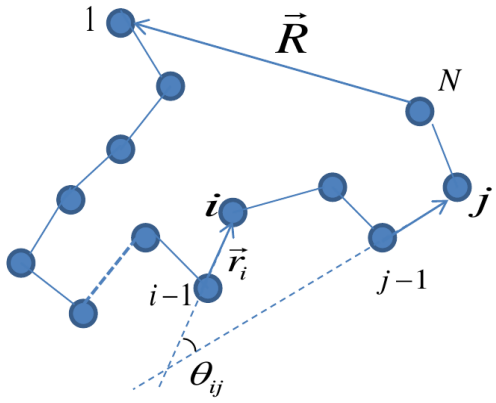


Fig. 1.2 The ideal chain consist of N monomers with bond vector $\vec{r}_i$ where $|\vec{r}_i| = a$. The end-to-end distance of the chain is $\vec{R}$, and θ_{ij} is the angle between the ith and jth bond vectors.

Distribution of end-to-end distances

Here, we study the end-to-end distance distribution of an ideal chain. Considering that every possible conformation of an ideal chain can be mapped onto a random walk, in which the number of walk steps is equal to the number of bonds of the chain [15]. In the freely jointed chain, the length of the bonds are constant and the orientation of each bond vector is independent of the other bond vectors [1, 2]. Therefore, every conformation of an ideal chain can be mapped onto the trajectory of random walk of a particle. We examine a symmetric one dimensional random walk given N total steps. If we have n positive (right) steps and $N-n$ negative (left) steps with length a, then the net displacement is $x = a(2n-N)$, and the probability of such a displacement is given by:

$$P_n(N) = \frac{\left(\frac{1}{2}\right)^N N!}{n!(N-n)!} \tag{1.3}$$

Since $a = \sqrt{\frac{<x^2>}{N}}$ and subsequently $n = \frac{N}{2} + \frac{x\sqrt{N}}{2\sqrt{<x^2>}}$, we have:

$$P_n(N) = \frac{\left(\frac{1}{2}\right)^N N!}{[\frac{N}{2} + \frac{x\sqrt{N}}{2\sqrt{<x^2>}}]![\frac{N}{2} - \frac{x\sqrt{N}}{2\sqrt{<x^2>}}]!} \tag{1.4}$$

The displacement of the particle can be treated as a continuous variable when the number of steps is large. By considering the limit $P_n(N)dn \to P(x)dx$, one can rewrite the probability density:

$$P(x) = lim_{N\to\infty}\left[\frac{\sqrt{N}}{2\sqrt{<x^2>}}P_n(N)\right] \tag{1.5}$$

$$P(x) = lim_{N\to\infty}\left[\frac{\frac{\sqrt{N}}{2\sqrt{<x^2>}}\left(\frac{1}{2}\right)^N N!}{[\frac{N}{2} + \frac{x\sqrt{N}}{2\sqrt{<x^2>}}]![\frac{N}{2} - \frac{x\sqrt{N}}{2\sqrt{<x^2>}}]!}\right] \tag{1.6}$$

In order to evaluate the probability density $P(x)$ for large value of N, we use Stirling's approximation for factorial function $n! \to n^n e^{-n}$. After some manipulations, the probability density can be written as:

$$\begin{aligned} P(x) = lim_{N\to\infty}\frac{1}{\sqrt{2\pi <x^2>}}\frac{1}{\sqrt{1-\frac{x^2}{N<x^2>}}}\left(1+\frac{x}{\sqrt{N<x^2>}}\right)^{-\frac{N}{2}-\frac{x\sqrt{N}}{2\sqrt{<x^2>}}} \\ \times\left(1-\frac{x}{\sqrt{N<x^2>}}\right)^{-\frac{N}{2}+\frac{x\sqrt{N}}{2\sqrt{<x^2>}}} \end{aligned} \tag{1.7}$$

$$P(x) = lim_{N\to\infty} \frac{1}{\sqrt{2\pi < x^2 >}} \left(1 - \frac{x^2}{N < x^2 >}\right)^{-\frac{N}{2}-\frac{1}{2}} \left(1 + \frac{x}{\sqrt{N < x^2 >}}\right)^{-\frac{x\sqrt{N}}{2\sqrt{<x^2>}}} \times \left(1 - \frac{x}{\sqrt{N < x^2 >}}\right)^{\frac{x\sqrt{N}}{2\sqrt{<x^2>}}} \tag{1.8}$$

We can further simplify the probability density function by using the limiting definition of the exponential $e^b = lim_{n\to\infty}\left(1+\frac{b}{n}\right)^n$. By using this limit, the two last terms cancel each other, and we have:

$$P(x) = \frac{1}{\sqrt{2\pi < x^2 >}} exp\left(\frac{-x^2}{2 < x^2 >}\right) \tag{1.9}$$

Considering that $< x^2 >=< y^2 >=< z^2 >= \frac{Na^2}{3}$ and $P(\vec{R}) = P(R_x)P(R_y)P(R_z)$, the probability distribution function of the end-to-end distance of an ideal chain in three dimension can be written as:

$$P(\vec{R}) = \left(\frac{3}{2\pi Na^2}\right)^{\frac{3}{2}} exp\left(-\frac{3(R_x^2 + R_y^2 + R_z^2)}{2Na^2}\right) \tag{1.10}$$

$$= \left(\frac{3}{2\pi Na^2}\right)^{\frac{3}{2}} exp\left(-\frac{3\vec{R}^2}{2Na^2}\right) \tag{1.11}$$

Freely-rotating chain model

In the freely-rotating model of polymer chains the torsion angle between monomers can freely adapt any values [1]. Essentially, the probability distribution of the torsion angles is distributed equally and independently between $-\pi < \phi < \pi$ for each monomer. In this model, both the bond length (a), and the bond angle (θ) are fixed. In order to calculate the mean square end-to-end distance (Eq. 1.3) of the chain, the correlation between bond vectors must be determined. The correlation is defined as the projection of the $\vec{r}_j$ vector on the $\vec{r}_{j-1}$ vector, that is equal to $a\cos(\theta)$.

Bond vector $\vec{r}_{j-1}$ passes the correlation between monomer jth and $(j-2)th$ to $a(\cos\theta)^2$, so that the correlation of ith and jth bond vectors will be equal to $a^2(\cos\theta)^{|i-j|}$. Given these assumptions the mean square of end-to-end distance of a freely-rotating chain can be written as:

$$
\begin{aligned}
<\vec{R}^2> &= \sum_{i=1}^{N-1}\sum_{j=1}^{N-1} <\vec{r}_i.\vec{r}_j> \\
&= \sum_{i=1}^{N-1}\left(\sum_{j=1}^{i-1} <\vec{r}_i.\vec{r}_j> + <\vec{r}_i.\vec{r}_i> + \sum_{j=i+1}^{N-1} <\vec{r}_i.\vec{r}_j>\right) && (1.12)\\
&= \sum_{i=1}^{N-1} <\vec{r}_i.\vec{r}_i> + a^2\sum_{i=1}^{N-1}\left(\sum_{j=1}^{i-1}(\cos\theta)^{|i-j|} + \sum_{j=i+1}^{N-1}(\cos\theta)^{|j-i|}\right) && (1.13)\\
&= a^2(N-1) + a^2\sum_{i=1}^{N-1}\left(\sum_{k=1}^{i-1}(\cos\theta)^k + \sum_{k=1}^{N-1-i}(\cos\theta)^k\right) && (1.14)
\end{aligned}
$$

Supposing that the decay of $(\cos\theta)^k$ is fast enough we can use the following approximation:

$$
\begin{aligned}
\sum_{i=1}^{N-1}\left(\sum_{k=1}^{i-1}(\cos\theta)^k + \sum_{k=1}^{N-1-i}(\cos\theta)^k\right) &\simeq 2\sum_{i=1}^{N-1}\left(\sum_{k=1}^{\infty}(\cos\theta)^k\right) && (1.15)\\
&= 2(N-1)\sum_{k=1}^{\infty}(\cos\theta)^k = 2(N-1)\frac{\cos\theta}{1-\cos\theta} && (1.16)
\end{aligned}
$$

$$
<\vec{R}^2> = a^2(N-1)\frac{1+\cos\theta}{1-\cos\theta} \qquad (1.17)
$$

Eq. 1.17, shows that the mean square end-to-end distance of the freely rotating chain is a simple function of the number of bonds $N-1$, the length of each bond a, and the fixed bond angle θ.

Worm-like chain model

The worm-like chain, or Kratky-Porod chain, is similar to the freely-rotating model but with very small bond angles [1]. This mean that the chain is fairly straight and stiff, that makes this model a suitable choice for studying rod like chains. The mean square end-to-end distance of the worm like chain can be approximated by using the exponential decay of the correlation coefficient along the chain:

$$< \vec{R}^2 >= a^2 \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} < \cos\theta_{ij} >= a^2 \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} exp\left(\frac{-|j-i|}{l_p} a\right) \tag{1.18}$$

where $l_p = \frac{-a}{ln(\cos\theta)}$. The summation over all monomers can be changed into an integration over the total length of the chain, so $a\sum_{i=1}^{N-1} \rightarrow \int_0^{R_{max}} du$ and $a\sum_{j=1}^{N-1} \rightarrow \int_0^{R_{max}} dv$.

$$< \vec{R}^2 > = \int_0^{R_{max}} \left[\int_0^{R_{max}} exp\left(\frac{-|u-v|}{l_p}\right) dv \right] du \tag{1.19}$$

$$= 2l_p R_{max} - 2l_p^2 \left(1 - exp\left(\frac{-R_{max}}{l_p}\right)\right) \tag{1.20}$$

In the rod like limit the persistence length is much longer than the contour length of the chain $l_p >> R_{max}$, therefore the exponential in the Eq. 1.20 can be expanded as follow:

$$exp\left(\frac{-R_{max}}{l_p}\right) \simeq 1 - \frac{R_{max}}{l_p} + \frac{1}{2}\left(\frac{R_{max}}{l_p}\right)^2 + .. \tag{1.21}$$

By substituting the expanded expression in the Eq. 1.20 the end-to-end distance of the rod like chain can be written as:

$$< \vec{R}^2 > \simeq R_{max}^2 \tag{1.22}$$

In the limit that the chain is flexible $R_{max} >> l_p$ the end-to-end distance reduces to $\simeq 2 l_p R_{max}$. Due to the bending modes, the end-to-end distance of a worm-like chain varies between these two limited values. The important difference between a worm-like chain and a freely jointed chain is that the former is flexible in the length scales shorter than the Kuhn length of the chain l_p, while the latter is assumed to be completely rigid on the length scales shorter than its Kuhn length "a" [1].

Hindered rotation model

The hindered rotation model also assumes a constant bond length and angle [1]. In this model the torsion angles are independently taken to be hindered by potential $U(\phi_i)$, such that each microstates probability will be proportional to the Boltzmann factor $exp\left(\frac{-U(\phi_i)}{k_B T}\right)$.

Depending on the functionality of the torsion potential, the hindered rotation model predicts the following end-to-end distance for the polymer chain:

$$< \vec{R}^2 >= a^2 (N-1) \left(\frac{1+\cos\theta}{1-\cos\theta}\right) \left(\frac{1+ < \cos\phi >}{1- < \cos\phi >}\right) \tag{1.23}$$

where $< \cos\phi >$ is the average *cosine* of the torsion angle that is populated by Boltzmann factor:

$$< \cos\phi >= \frac{\int_0^{2\pi} \cos\phi \, exp\left(\frac{-U(\phi)}{k_B T}\right) d\phi}{\int_0^{2\pi} exp\left(\frac{-U(\phi)}{k_B T}\right) d\phi} \tag{1.24}$$

Free energy of ideal chain

The occurrence of a microstate of an ideal chain with a specific end-to-end distance is proportional to the all microstates with that specific end-to-end distance over all microstates for any value of R between 0 and R_{max}. Considering the non-correlated orientation in the bond vectors of the ideal chain, one can apply the statistics of the random walk in the statistics of the ideal chains [1]. As already proved, the end-to-end distance of such an ideal chain in $3D$ can be written as:

$$P(N,R) = \left(\frac{3}{2\pi Na^2}\right)^{\frac{3}{2}} exp\left[-\frac{3R^2}{2Na^2}\right] \tag{1.25}$$

where N is the number of monomers. By using Eq. 1.25 and Eq. 1.26, the Helmholtz free energy of the ideal chain can be derived explicitly:

$$F(N,R) = -k_B T ln P(N,R) \tag{1.26}$$

$$F(N,R) = -k_B T ln\left(\left(\frac{3}{2\pi Na^2}\right)^{\frac{3}{2}} exp\left[-\frac{3R^2}{2Na^2}\right]\right) \tag{1.27}$$

$$= \frac{3k_B TR^2}{2Na^2} - \frac{3}{2}k_B T ln\left(\frac{3}{2N\pi a^2}\right) \tag{1.28}$$

$$= \frac{3k_B TR^2}{2Na^2} + F(N,0) \tag{1.29}$$

As shown in Eq. 1.29, the free energy of an ideal chain with end-to-end distance R is similar to the energy of a $1D$ simple harmonic oscillator with relaxation length and stiffness of 0 and $\frac{3k_B T}{Na^2}$, respectively.

1.2.2 Real Chain: Excluded volume and Flory theory

So far we have studied ideal chains, with this particular feature that nonbonded interactions between monomers of the polymer are negligible. However, in polymeric systems with low concentrations of ions in solvent, the monomers of the chain that are not close in sequence can also interact and affect the dynamic of the polymer [1]. The nonbonded interactions, essentially drive the chain to show a variety of conformational features, hence, it is crucial to characterize them. Here we introduce a well established method to accomplish the presence of nonbonded interactions such as excluded volume interaction between monomers to study the physics of real polymer chains.

Excluded volume$\rightarrow$ In order to bring two monomers from infinity to within distance r in a solvent, one have to pay $U(r)$ energy that includes the steric repulsive interaction of two overlapping beads. We assume that the solvent is athermal, which implies that the molecules of the solvent interact with each monomer of the chain by same potential. We use Boltzmann factor to show the probability of finding two monomers within distance r at temperature T. When the monomers are fairly close, the interaction energy between the overlapped beads is positive and strong, that is similar to hard core repulsion interaction. Thus, it is very unlikely to find two hardcore monomers within short distance from each other.

The excluded volume of each monomer is defined as:

$$v = \int_0^\infty \left[1 - \exp\left(\frac{-U(r)}{k_B T}\right)\right] d^3 r \tag{1.30}$$

this quantity is the consequence of nonbonded interactions among monomers. Purely repulsive interactions increases the excluded volume, although attractive interactions between monomers entails to decrease in the excluded volume.

Flory theory$\rightarrow$ The real polymer chains in athermal and good solvent can quantitatively be studied by probing the competition between the entropy and the excluded volume interactions that tend to inflate the chain and reduce the number of available conformations [1]. By making use of Flory theory that consider no correlation between monomers, one can investigate in a simple way, the balance

between the entropy and energy. Assume that we have a chain with N monomers, the probability to find another monomer in the excluded volume of given monomer (v), is the product of excluded volume v and the density of the monomers in the pervaded volume of the chain $\frac{N}{R^3}$. Where $R > R_0 = aN^{\frac{1}{2}}$ is the size of the swollen polymer. The energetic cost of being excluded from this volume is k_BT per exclusion or $k_BT\frac{vN}{R^3}$ per monomer. If we have N monomer in our system then the total energy due to the excluded volume interaction is:

$$F_{int} \approx k_BTv\frac{N^2}{R^3} \tag{1.31}$$

In addition to the exclusive interaction between monomers, the entropic cost of the real chain should be added to the total energy. Which basically is the energy that is required to stretch the chain to the end-to-end distance of R. In Flory estimation the total free energy of the chain can be written as:

$$F_{tot} = F_{int} + F_{ent} \approx k_BT\left(v\frac{N^2}{R^3} + \frac{R^2}{Na^2}\right) \tag{1.32}$$

The end-to-end distance of the polymer at equilibrium can be obtained by minimizing the total energy of the chain ($\frac{\partial F_{tot}}{\partial R} = 0$), that give rise to $R_F \approx v^{\frac{1}{5}}a^{\frac{2}{5}}N^{\frac{3}{5}}$. If the excluded volume interaction energy is smaller than k_BT, the chain behave mainly similar to an ideal chain. Although by increasing the strength of excluded volume interaction, the chain inflate and swell up.

1.2.3 Real and Ideal chains under tension

Here, we study the behavior of an ideal and a real polymer chain under external tension. The end-to-end distance of an ideal and real chain without any external force, is given by:

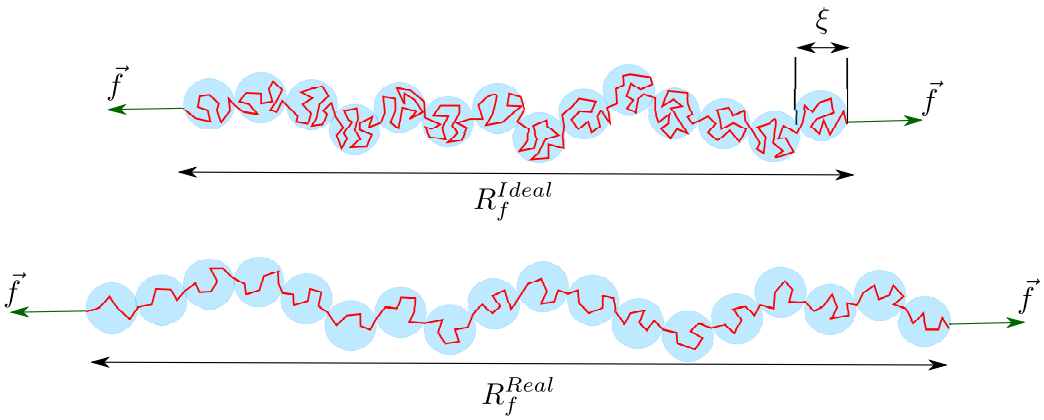


Fig. 1.3 In the top and bottom panels we show the ideal and real chains with same counter length that are under tension $\vec{f}$ at the ends, respectively. ξ is the diameter of the tension blobs, R_f^{Real} and R_f^{Ideal} are the end-to-end distance of the real and ideal chains respectively.

$$R_0^{Ideal} \simeq aN^{\frac{1}{2}} \tag{1.33}$$

$$R_0^{Real} \simeq aN^{\frac{3}{5}} \tag{1.34}$$

where a is the length of the fixed bonds and N is the total number of monomers of the chain.

By exerting an external force at the ends of the polymer chain, the monomers only manifest the effect of the constraints above a certain length scale that is know as tension blob. In the length scale smaller than the size of the tension blob, there is no impact on the monomer's motion due to the presence of external force at the ends of the chain. Since both ideal and real polymer chains are self-similar fractals, same scaling of the end-to-end distance can be applied to the end-to-end distance of that portion of the polymer chain that form the tension blob [1]. Thus, the size of the tension blobs for the ideal and real polymer chains can be written as:

$$\xi^{Ideal} \simeq a g^{\frac{1}{2}} \tag{1.35}$$

$$\xi^{Real} \simeq a g^{\frac{3}{5}} \tag{1.36}$$

where g is the number of monomers inside the tension blobs.

One can derive the end-to-end distance of the ideal and real chains under tension as follow:

$$R_f^{Ideal} \simeq \xi \frac{N}{g} \simeq \frac{N a^2}{\xi} \tag{1.37}$$

$$R_f^{Real} \simeq \xi \frac{N}{g} \simeq \frac{N a^{\frac{5}{3}}}{\xi^{\frac{2}{3}}} \tag{1.38}$$

The incurred free energy to stretching the chains is order of $k_B T$ per each blob:

$$F^{Ideal} \simeq k_B T \frac{N}{g} \simeq k_B T \left(\frac{R_f^{Ideal}}{R_0^{Ideal}} \right)^2 \tag{1.39}$$

$$F^{Real} \simeq k_B T \frac{N}{g} \simeq k_B T \left(\frac{R_f^{Real}}{R_0^{Real}} \right)^{\frac{5}{2}} \tag{1.40}$$

the corresponding stretching force for each polymer can be obtained by free energy differentiation with respect to end-to-end distance:

$$f^{Ideal} \simeq \frac{k_B T}{R_0^{Ideal}} \frac{R_f^{Ideal}}{R_0^{Ideal}} \tag{1.41}$$

$$f^{Real} \simeq \frac{k_B T}{R_0^{Real}} \left(\frac{R_f^{Real}}{R_0^{Real}} \right)^{\frac{3}{2}} \tag{1.42}$$

The force that is required to stretch a real polymer chain grows rapidly with its end-to-end distance, although it always is smaller than the corresponding force in the ideal chain. The reason for this discrepancy lies in the fact that stretching an ideal chain remarkably decreases the number of available configurations, however in the real chain, the presence of excluded volume interaction have already reduced the number of available configurations.

1.3 Rouse, Zimm, and Reptation dynamics

1.3.1 Rouse model

Rouse proposed the first successful realistic model of polymer chain dynamics in solvent [1]. In the Rouse model the polymer chain is modeled as N beads that are bonded by simple harmonic spring that its relaxation length is a. In this model the beads can interact with each other only through the bonded potential. The solvent assumed to drain through the chain as it moves, although the beads of the chain can feel a dragging force from the solvent that is proportional to the friction coefficient and the velocity. Considering that the beads of the polymer chain can interact independently with the solvent, the total friction force on the chain can be computed by summing over all friction forces on the beads, $F_{friction} = -N\zeta\vec{v}$, where ζ is the friction coefficient and $\vec{v}$ is the velocity of the chain.

Due to the presence of thermal fluctuations in the system the chain can diffuse. The diffusion coefficient of the polymer can be obtained from Einstein relation:

$$D_R = \frac{k_B T}{N\zeta} \tag{1.43}$$

The polymer may exhibit peculiar viscoelastic properties in the time scales that are smaller than the time that the polymer needed to diffuse in the distance of its own size, this characteristic time scale τ_R is known as Rouse time:

$$\tau_R \simeq \frac{R^2}{D_R} \simeq \frac{N\zeta}{k_B T} R^2 \tag{1.44}$$

In the time scales lager than the Rouse time, viscoelastic features of the chain disappear [1]. Therefore, the chain relieve the viscosity and there would not be any memory in the motion of polymer chain due to the friction force. If we scale the size of the chain by its end-to-end distance then the Rouse time can be rewritten as:

$$\tau_R \simeq \frac{a^2 \zeta}{k_B T} N^{1+2\nu} = \tau_0 N^{1+2\nu} \tag{1.45}$$

where τ_0 is the relaxation time of a free monomer.

1.3.2 Zimm model

As already mentioned, in the Rouse model of polymer chains, we assume that surrounding solvent of a monomer can not exert any additional force on the other monomers of the chain. However, in realistic systems when a monomer of polymer interact with the solvent molecules, it provides a velocity field that influences the motion of the other monomers of the chain. This effective interactions between the monomers of the polymer due to the presence of the solvent is know as hydrodynamic interaction that just only decays slowly with the distance between the monomers, $(\frac{1}{r})$ [1].

When the polymer chain drags the solvent into its pervaded volume, it moves approximately as a solid object of size $R \simeq aN^{\nu}$. The friction coefficient of a chain of size R in a solvent with viscosity η_s is given by:

$$\zeta_Z = R\eta_s \tag{1.46}$$

The diffusion coefficient of the polymer chain in Zimm model can be obtained by Einstein relation:

$$D_Z = \frac{k_B T}{\zeta_Z} \simeq \frac{k_B T}{R\eta_s} \simeq \frac{k_B T}{aN^{\nu}\eta_s} \tag{1.47}$$

The polymer chain in the Zimm model can diffuse as an individual particle with volume proportional to its pervaded volume in the solvent. During the Zimm time τ_Z, the polymer chain can diffuse a distance of its size:

$$\tau_Z \simeq \frac{R^2}{D_Z} \simeq \frac{\eta_s}{k_B T} R^3 \simeq \frac{\eta_s a^3}{k_B T} N^{3\nu} \simeq \tau_0 N^{3\nu} \tag{1.48}$$

By comparing the Rouse and Zimm characteristic times, we find out that the latter is smaller than the former. In Zimm model the chain feel lower friction force, therefore, the faster process occurs during Zimm motion. In the Rouse model the solvent can pass through the chain and exerts higher friction force on the polymer chain, however in the Zimm model, the chain motion is roughly similar to a solid particle with size of R.

1.3.3 Reptation in polymers

In melt and dense polymeric systems, polymers may share their pervaded volume [1]. Understanding the dynamic of dense polymers with this particular feature

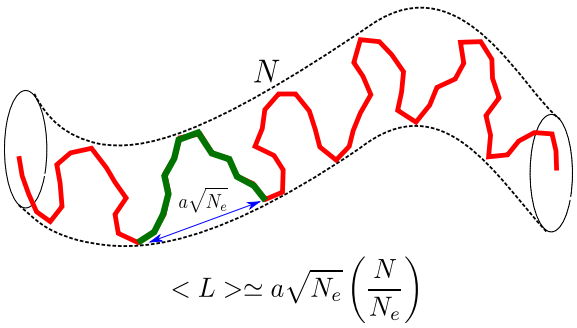


Fig. 1.4 Here we show a schematic view of the reptation model. The tube that is imposed by surrounding chains is depicted by dotted curves, the red curve shows the chain itself and the green part of the chain represents the loop that is created during entanglement between the chain and the surrounding chains. $< L >$ is the average counter length of the tube, N_e is the number of beads in the loop and N is the number of the chain beads.

is a difficult task. In such a melt system, the dynamic of polymers appears to be similar to many body problem. However, by using Edwards tube concept one can reduce the complexity of the problem to the motion of a single chain that is confined into a tube which is formed by its surrounding chains. The simplest tube model that is known as reptation model, is proposed by de Gennes in order to study the dynamic of the linear entangled polymers [1]. In this model, the entangled polymer chain moves along its confining tube due to the simple diffusion of the small loops of the chain (that are formed during the polymer entanglement with the surrounding chains) along the counter of the pervaded volume or permeative path. The diffusion coefficient D_c of the chain along the tube is indeed the Rouse diffusion coefficient:

$$D_c \simeq \frac{k_B T}{N\zeta} \tag{1.49}$$

The reptation time is the time that chain required to diffuse out of a tube with average length $< L >$:

$$\tau_{rep} \simeq \frac{< L >^2}{D_c} \simeq \frac{\zeta a^2}{k_B T} \frac{N^3}{N_e} = \frac{\zeta a^2}{k_B T} N_e^2 \left(\frac{N}{N_e} \right)^3 \tag{1.50}$$

here we used $< L > \simeq \frac{aN}{\sqrt{N_e}}$ for the counter length of the tube, where N_e is the number of Kuhn monomers in the loops.

The first part of the last term in the reptation time is simply the Rouse time of a chain containing N_e monomers:

$$\tau_e \simeq \frac{\zeta a^2}{k_B T} N_e^2 \tag{1.51}$$

The ratio of the reptation time and τ_e is the cube of the number of entanglements along the chain:

$$\frac{\tau_{rep}}{\tau_e} \simeq \left(\frac{N}{N_e} \right)^3 \tag{1.52}$$

In the reputation time, the chain diffuses a distance of its own size R:

$$D_{rep} \simeq \frac{R^2}{\tau_{rep}} \simeq \frac{k_B T}{\zeta} \frac{N_e}{N^2} \tag{1.53}$$

1.4 Introduction to topology

Topology in general is a fascinating topic that has captured scientists attention for more than two centuries and has progressed remarkably from its early rise. Although the core question of how effectively classify knots and links is as elusive as before [10–13]. Knot are complex 3D objects that mathematically are defined

in closed arcs [10]. Sufficiently long strings such as ropes and whatever looks like it can be knotted and exhibit a variety of features that are knot specific [12].

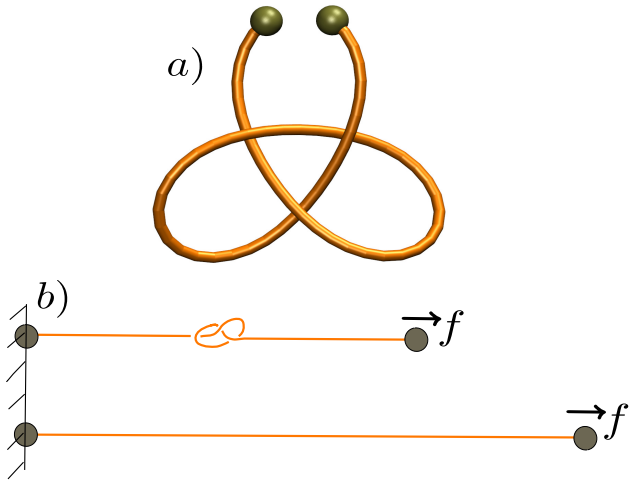


Fig. 1.5 (a) We show a trefoil knot on a rubber. Closing the ends of the chain that are depicted as dark green spheres will fix the topology. (b) Top and bottom strings are constrained knotted and unknotted chains respectively. Grafting one end of the rubbers and pulling the other end by same force f, will accomplish a certain end-to-end distance in the rubbers. In particular the rubber that contains a knot, due to its self-entanglement exhibit smaller end-to-end distance in comparison to a unknotted rubber.

Although every knot on a string has irregular conformation, topology of stiff chains enforce almost a well formed geometrical structure; i.e. the geometry of the knot can saliently be evident for simple knots on stiff rubbers, but as the rubber gets more flexible the clarity of its structure will be lost. In Fig. 1.5 *a*, a trefoil knot on a rubber is shown, by closing the two ends of the rubber that are depicted as dark green spheres, the topology will be fixed. The latter implies that changing the geometry of the closed rubber, would not affect its topology [11]. A very crude way to find out if a rubber is self-entangled or knotted is to pull by its ends. Under certain external tension $\vec{f}$, the end-to-end distance of the knotted rubber would be smaller than the other rubber that does not have any knot for sure. In Fig. 1.5*b*, an illustration of the scheme is shown. The ratio of the end-to-end distance of the knotted and unknotted rubber depends on the complexity of the embedded knot.

Here, we would like to provide a basic introduction to knot theory, and discus a few examples of biological systems where knots can occur and affect the function of the system.

1.5 Topological invariants

To address the question of how the topology of a knot or entangled rubbers can be determined, we need to study the methods that one can employ to identify and classify the topological state of linked or knotted rubbers. Here, we aim to introduce some topological invariants of entangled chains.

1.5.1 Gaussian link number

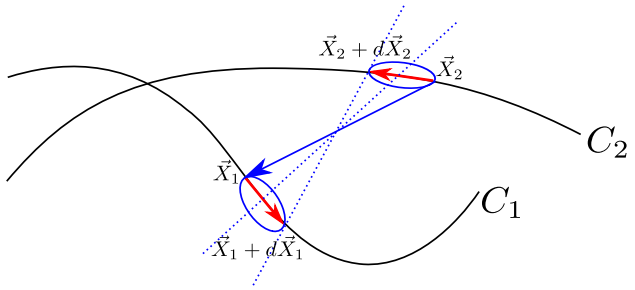


Fig. 1.6 C_1 and C_2 represent the intersected parts of simple closed curves whose elements are $d\vec{X}_1$ and $d\vec{X}_2$.

In 1833, Gauss wanted to calculate the work that is required to be done on a magnetic pole for moving it along a closed curve in the presence of a loop of current. During his progress in the processing he discovered what is known as Gauss Linking Number [11, 12].

In order to calculate the Gauss Linking Number of two intersected rings, we consider two simple closed curves, C_1 and C_2. The Gaussian linking number of entangled circles is an isotopy invariant that is defined as:

$$L_{12} = \frac{1}{4\pi} \oint_{C_1} \oint_{C_2} \frac{(\vec{X}_1 - \vec{X}_2).d\vec{X}_1 \times d\vec{X}_2}{|\vec{X}_1 - \vec{X}_2|^3} \tag{1.54}$$

if we label $\vec{r} = \vec{X}_1 - \vec{X}_2$ then:

$$L_{12} = \oint_{C_1} \oint_{C_2} \frac{\vec{r}.d\vec{X}_1 \times d\vec{X}_2}{4\pi r^3} \tag{1.55}$$

we consider the apparent intersection between an element dX_1 of C_1 and an element dX_2 of C_2 on a sphere whose radius is the distance between the segment of the curves, thus the solid angle of the intersect of the two elements will be given by:

$$d\omega = 2\frac{\hat{r}.d\vec{X}_1 \times d\vec{X}_2}{4\pi r^2} \tag{1.56}$$

since by integrating once around C_1 and once around C_2, we contribute twice to the solid angle, thus by assigning $+$ and $-$ to each apparent intersection that are referring to over and under-crossings respectively, we can rewrite the Gaussian link number as follow:

$$L_{12} = \frac{1}{2}\int_{\Omega} d\omega = \frac{1}{2}\sum_{r \in C_1 C_2} \varepsilon_r; \quad \varepsilon_r = \pm 1 \tag{1.57}$$

The minimum number of crossings c_{min} over all possible projections of entangled circles is a topological crossing number of the system and it is an invariant. In general for N curves in space the linking number can be written as: $L_{ij} = L(C_i, C_j); i \neq j$, hence for N component link, the Gaussian liking number given by:

$$L = \sum_{i \neq j}^{N} L_{ij} \tag{1.58}$$

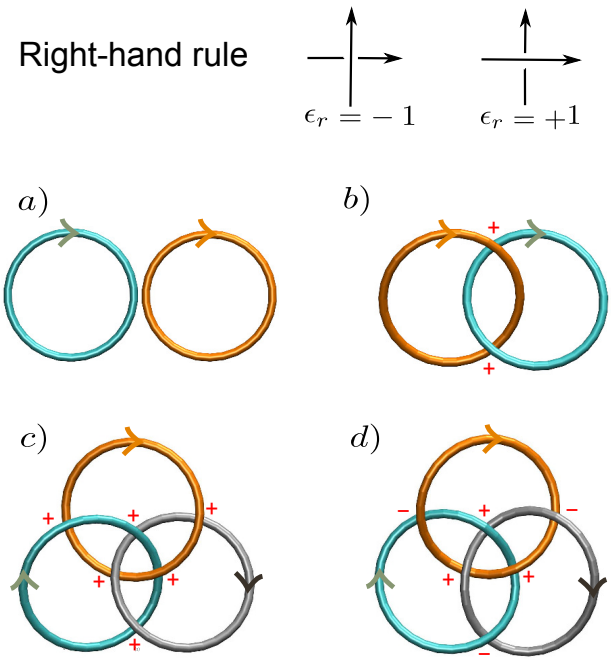


Fig. 1.7 In figure a, b, c and d, we show two unlinked circles, two linked circles, three Hopf-linked circles and the Borromean rings that consists of three entwined rings, respectively. At each panel we show the crossing pattern of the circles by making use of the right-hand rule that is depicted in the top of figure.

In Fig.1.7, we show some linked rings with different c_{min} and L:

- a) Two separated rings: $c_{min} = 0,\ L(C_1, C_2) = 0$
- b) Two Hopf-linked rings: $c_{min} = 2,\ L(C_1, C_2) = +1$
- c) Three Hopf-linked rings: $c_{min} = 6,\ L(C_1, C_2, C_3) = +3$
- d) Borromean rings: $c_{min} = 6,\ L(C_1, C_2, C_3) = 0$

1.5.2 Călugăreanu-White invariant

Gauss's works motivated scientists to pursue links and knots in different fields and employ them to understand the physical properties of the relevant systems. Particularly in biology, because of the specific structure of DNA, it can be difficult to understand the conformational features that are arising from topology, without having clear insight of the underlying mechanisms that links DNA geometry to its

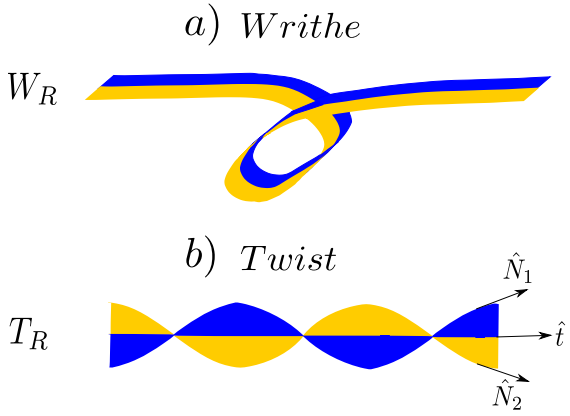


Fig. 1.8 In panel a we show a writhed ribbon that is illustrated by blue and yellow counterparts. In panel b the ribbon is twisted, $\hat{t}$, $\hat{N}_1$, $\hat{N}_2$ are the tangential vector on symmetry axis and the tangential vector on the edges of the blue and yellow counterparts, respectively.

topology [10, 12]. One can study the physics of DNA by making use of a Ribbon-model that is similar to the conformation of DNA. We facilitate the understanding of DNA topology by considering two counterpart ribbons that represent the backbones of double strand DNA, which are colored as blue and yellow in Fig. 1.8.

Writhe and twist are geometric quantities that one can use to characterize the conformational properties of a ribbon modeled DNA molecule.

Writhe: We label the the two edges of the ribbon in Fig. 1.8a, by : $C_1 : \vec{X}_1(s)$ and $C_2 : \vec{X}_2(s)$, where s is the arc length of the ribbon. The writhing number W_R of a single closed curve is defined by:

$$W_R = \frac{1}{4\pi} \oint_C \oint_C \frac{(\vec{X} - \vec{X}^*).d\vec{X} \times d\vec{X}^*}{|\vec{X} - \vec{X}^*|^3} \tag{1.59}$$

where $\vec{X}$ is the vector coordinate to the symmetry axis of the Ribbon.

Twist: The total twist T_R of the ribbon $R_C(\hat{t})$ is defined by:

$$T_R = \frac{1}{2}\oint_C (\hat{N}_1 \times \hat{N}_2).\hat{t}ds = \tau(C) + \Pi \tag{1.60}$$

where $\tau(C)$ is the total torsion and Π is the intrinsic twist of the ribbon.

The Călugăreanu-White linking number L of a ribbon $R_C(\hat{t})$ is the sum of the writhe and twist numbers:

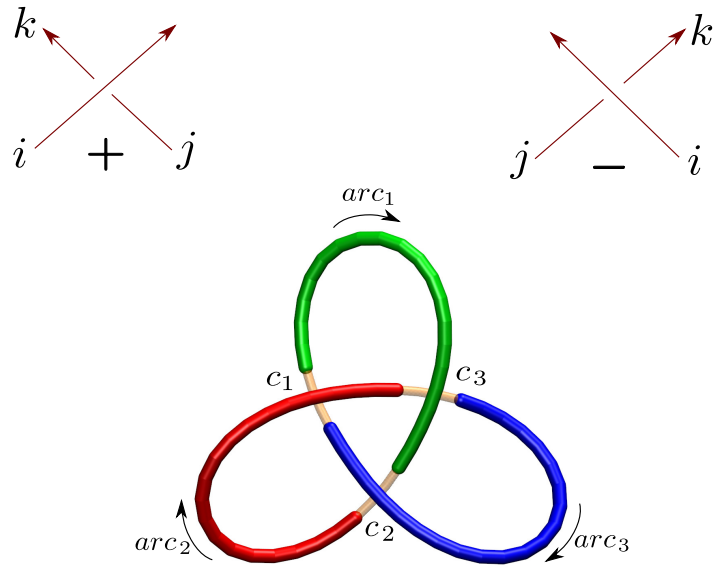
$$L = W_R + T_R \tag{1.61}$$

that is an isotopy invariant of the ribbon [11, 13]. The Călugăreanu-White linking number is an integer and as topological invariant does not depend on the continuous deformation of the ribbon. Considering the fact that twist and writhe are not a well defined intuitive concepts, what is remarkable about Călugăreanu-White linking number is that the sum of these geometric quantities is an invariant topological property.

1.5.3 The Alexander Polynominal

Several topological invariants such as knot groups, Jones, and Alexander Polynomial are introduced to distinguish different type of knots.

The determination of the knot invariant require the projection of the three dimensional conformation of the knot onto a plane, that is known as knot diagram [10, 12, 13]. It is demanded that a knot diagram contain the information of the over and under crossings, such that one can reconstruct the original knot from its diagram [10]. As an example, the diagram of knot 3_1 with minimum number of crossings is shown in the center of Fig. 1.9, that one can easily calculate the number and the type of crossings. Assuming that t is the variable of the polynomial, the Alexander polynomial of a knot can be determined based on the following steps:



$$P(t) = \begin{array}{c} \begin{array}{ccc} arc_1 & arc_2 & arc_3 \end{array} \\ \begin{pmatrix} t & 1-t & -1 \\ -1 & t & 1-t \\ 1-t & -1 & t \end{pmatrix} \end{array} \begin{array}{c} c_1 \\ c_2 \\ c_3 \end{array} \quad , \qquad \Delta(t;3_1) = t^2 - t + 1$$

Fig. 1.9 In the top and center of figure, the right hand rule and the diagram of knot 3_1 are depicted, respectively. The green, red and blue colored portions on the knot show the three arcs separated by two consecutive under crossings.

1. Choose an arbitrarily orientation on the knot diagram and define the sign of each crossing based on the right-hand rule that is shown in top of Fig 1.9.

2. Take an arbitrary point of the knot diagram that is not on the crosses, follow the segment orientation and assign an increasing numbering index to all under-crosses.

3. Consider the separate arcs as going from an under-cross to the next under-cross and assign to each under-crosses arc a numbering index as shown in Fig 1.9.

4. Define an $n \times n$ matrix $P(t,\tau)$. The rows of the matrix correspond to the crossings in the knot diagram and the columns to the arcs. For each crossing we assume that from definition the arc i passes over arcs j and k. First, we

set all the elements of the matrix equal to zero, and construct the matrix of the knot diagram with the following values in the entries $P(c,i)$, $P(c,j)$ and $P(x,j)$ of all crossings.

- If c is positive crossing, $P(c,i) = 1-t$, $P(c,j) = -1$, $P(c,k) = t$.
- If c is negative crossing, $P(c,i) = 1-t$, $P(c,j) = t$, $P(c,k) = -1$.
- If $i = k$ or $i = j$, $P(c,j) = 1$ and $P(c,k) = -1$ in this condition the sign of crossing does not matters.

5. Compute any minor of order $n-1$ of the matrix knot diagram and multiple it by $\pm c^m$, $-m \in \mathbb{N}$, in such a way that the resulting polynomial does not contain negative powers, and has a positive constant elements. Such a minor is the (irreducible) Alexander polynomial $\Delta(t;\tau)$ of the knot.

As an instance, in Fig 1.9, we have computed the Alexander polynomial of knot 3_1. As already mentioned the Alexander polynomial is a topological invariant and does not depend on the geometry of the knot. The Alexander polynomial can distinguish the topological invariant of the knot, but it can not determine the chirality of the given topology, because swapping the positive and negative signs in the matrix of the knot diagram can not change the resulted Alexander polynomial.

Different knots with complicated topologies may have same Alexander polynomial, which makes the Alexander polynomial as a method that is not as powerful as others ways that one can exactly determine the topological invariant of the knots. Although, for the simple knots that have less than 11 crossings, only 6 topologies have same Alexander polynomial [10]. The advantages of the Alexander polynomial is that the required calculations and the computation cost for the related knot is very short and straightforward.

1.6 Classification of knots

1.6.1 knots identification: closure scheme

The study of topological features of knotted object requires its exact definition. Knot in mathematics is defined only for closed curves, however, in most physical

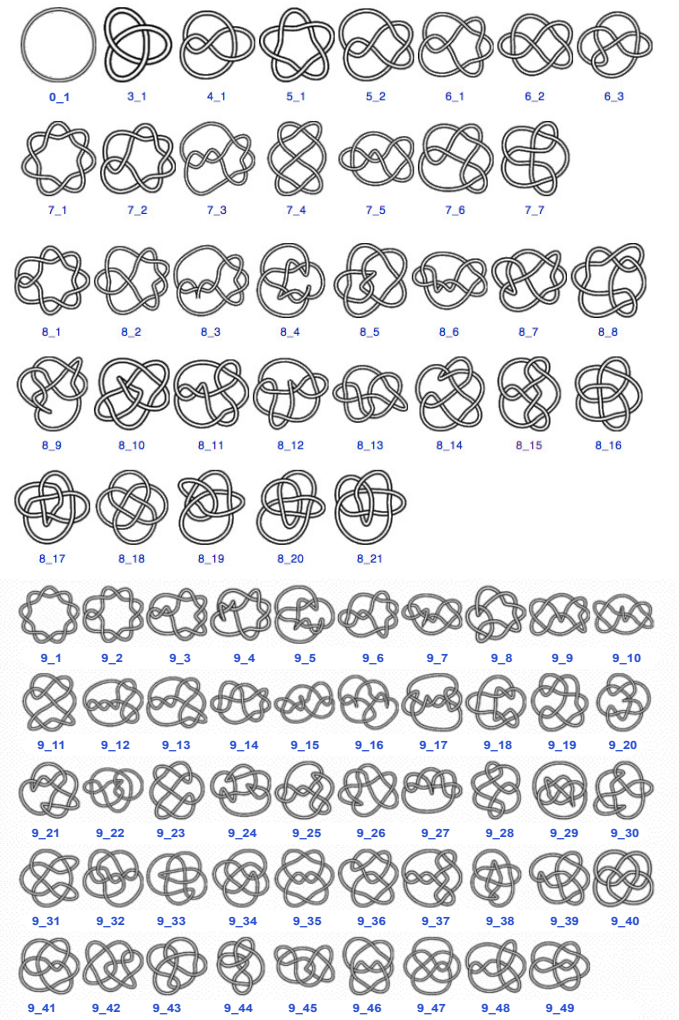


Fig. 1.10 All knots with number of crossing less than 10.

systems that might come under study the chains are open. Therefore, for identification of knot, it is crucial to find a way to closing the open knotted chains without affecting the topology by this operation.

Different methods are introduced to identify the knots in open chains. In our work we use a suitable closure scheme that based on this procedure the knot is characterized by prolongating the terminus of the knot and closing them with an arc in sufficiently far distance from the knot center, after the closure and circulation of the knotted chain, one can classify the complexity of the knot by employing Alexander Polynomial which as topological invariants does not depend on the

geometry of the closed chain [151, 152]. In Fig. 1.10 we show the knots tabulation whose number of crossing is less than 10.

1.6.2 Composite knots

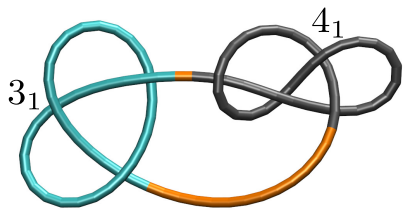


Fig. 1.11 Here we show an example of composite knot on a ring. The cyan part of the chain represent a trefoil knot and the black part is a 4_1 knot.

Multiple knots can occur in single polymer chain such as proteins and dsDNA. The knots on the polymer chain can interact and enforce specific structural and dynamical properties. No anti-knot exist; it is impossible to reduce the complexity of a knot by embedding another knot in the same chain. Composite knot consists of connected knots which are independent of each other, thus every composite knot can be decomposed into its prime component that are well defined knots. Composite knots are associative and also commutative. Multiple knots can locally influence polymers conformation such that it could vary from one knotted portion to other knotted part of the chain [10, 13, 151]. As an example of composite knots in Fig. 1.11, we have shown a doubly knotted polymer ring that consists of knot 3_1 and knot 4_1.

1.7 Topology in biopolymers

Topological constraints are ubiquitous in soft materials such as liquid crystals and biopolymers. Protein is linear polymer that consists of building blocks called amino acid [16, 17, 19–23]. The sequence of bases along the DNA determines the sequence of amino acids in proteins. There are 20 different amino acids in proteins, however only 4 distinguished bases in DNA are exist. Each amino acid is specified by a codon that is a group of three bases [7, 8]. Despite the fact that the sequence of codons on DNA determines the sequence of amino acids in proteins,

the DNA helix does not itself play a role in proteins synthesis. The translation of the sequence codon into amino acid occurs by the third class of molecule-messenger called RNAs [7]. During proteins, DNA and RNA transactions, the topological constraints can form and determine the characteristic features of the macromolecules and play pivotal role in their function in biological system. Here we illustrate a few examples of topologically constrained proteins and DNAs.

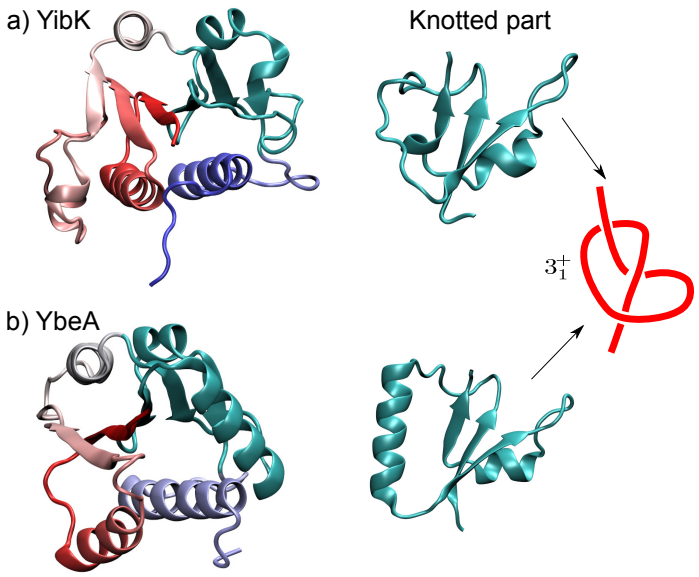


Fig. 1.12 In panel a and b, we show a cartoon representation of two simple knotted proteins, YibK from Haemophilus and YbeA from Escherichia coli, respectively. Both proteins host a right handed trefoil knot that is shown in the right side of each panel.

Knot in proteins→ The failed assumption that proteins folding avoids knots in order to prevent kinetic traps was held until 2000 [18]. However during the recent decade, advances in experimental tools reveals that proteins can form complicated tertiary structures such as knots. A broad interdisciplinary community spanning from theoretical physicists to experimental biologists have used their own approaches to understand the mechanism that drives the proteins to fold into knotted native state [16, 17, 19, 20]. In fact computational tools in many ways can complement the elucidated observations by experiments, however contrasts can be expected too.

Knotted proteins are an elegant example of biologically self constrained systems [36, 40, 44, 60, 61]. Experimentally it has been shown that knots in proteins occur in denatured state, and that the required denaturation rate to knot depends on their structural complexity [89]. Among all protein knots, some of them are deeply embedded into the polypeptide chain, while some of knots are shallow. The latter knotted proteins have included most part of the polypeptide chain within the knot and the topology is expanded, such that removing a few residues from the terminus of the proteins unravel the knot.

A unifying comprehension of the functional role played by knots in proteins has not found yet, although it is worth mentioning that topological constraints affect the structural features of proteins by encompassing their polypeptide chain which increases their thermodynamical and mechanical stability [89–92].

In fact, protein knotting is statistically and evolutionarily a rare event [89, 92]. There are notable differences between knotted proteins and other knotted biopolymers; the length, the location and the type of knot always remains specific for each knotted protein and also it is reproducible by repeating the folding process from chemically denatured unfolded conformation [90, 91].

In Fig. 1.12 panel a and b, we show two examples of knotted proteins, the YibK protein from Haemophilus and YbeA protein from Escherichia coli, respectively. Both proteins host a right handed trefoil knot, where close to 40 sequence of amino acids of their structure are similar (the knotted part of the proteins are shown in the right panel of each protein).

Recent observations by Mallam and Jackson in the experimental study of YibK and YbeA folding, provides crucial information regarding to the mechanisms of knotting in proteins [89]. They claimed that the mentioned proteins can spontaneously fold into their knotted native state without the aid of chaperones. This report, provides to us the clue that the internal interactions between the amino acids of the polypeptide chain might be enough to drive the chain into the knotted conformation.

Knot in DNA and RNA$\rightarrow$ DNA at its equilibrium state has a right-handed helical structure with a diameter of approximately $10A$. Every pitch of the helix includes 10 base pairs that giving to $34A$ for the length of the pitch [6, 8]. In human nucleus, 46 chromosomes form $2m$ of tangled DNA. The chromosomes deliberately tie themselves in knots and then untangled. During mitosis, topoisomerase cuts

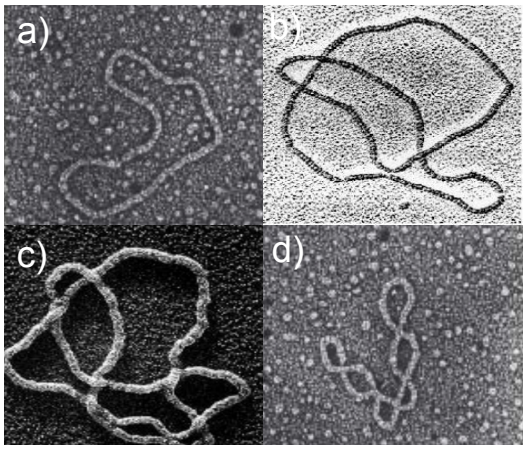


Fig. 1.13 Here, we show some experimentally captured image of circulated DNAs. In panel a, the DNA ring is unknotted, in panel c and d, the DNA ring host a 3_1 and 6_1 knot respectively, in panel d the DNA is knot free although it formed a coil [14, 7].

the entangled DNAs at any place where the helix is bent or under strain. When two chromosomes are tangled and exert stress on each other, topoisomerase cuts the entangled DNAs in such a way that one DNA can pass through without being hindered by the other DNA. During the passage, topoisomerase enzyme rejoins the cut ends once again [8].

Prokaryotic genomic DNAs and many viral DNAs are circular; i.e. without free ends. Knots in such a closed DNAs can form during replication, and transcription. Unwinding of the circular knotted DNA molecule induces strain on the other portion of DNA and form super coils. Topoisomerase are active in nucleus and relieve any DNAs strain due to their entanglement [8]. It is worth to mention that the release of the stress in the DNA super coil is a duty for topoisomerase I, that operates by cutting and rejoining one strand of DNA. However topoisomerase II relieve the DNA by cutting both strands of DNA. In Fig. 1.13, we show some photo of DNA rings experiment, in panel a, DNA is knot free, in panel b and c the DNA ring host a 3_1 and 6_1 knot respectively, in panel d the DNA formed a coil [14, 7].

The structure of single-strand RNA is generally similar to DNA strand. Most cellular RNAs are single strand and show a variety of conformations. Difference in the length and various configurations that RNAs can adopt, allow them to carry out many function in cells. The complementary bases of RNAs can stick to

each other and form loops that are simplest secondary structure in single strand RNAs [8, 7]. RNA is chemically active, that lead to many interactions between different segments of the chain. Recently, it has been shown that the knotting ability of RNA due to the presence of such a interaction along the chain is much lower than DNA [104].

1.8 Outline

We devote the rest of this Thesis to understand the structural and dynamical features of self-entangled and linked biopolymers such as proteins and DNAs. Thus, we address the following questions: How do knots form in proteins? How likely is DNA knot? What are characteristic properties of topologically constrained and linked DNAs?

In chapter 2, we study the folding mechanisms of two small knotted proteins. In general protein folding depends both on the intrinsic properties of the polypeptide chains and the effects of its environment. In fact, there are two main reasons that govern the proteins to fold into their native structure from a denatured conformation:

1. Interaction between the polypeptide chain and the surrounding
2. Interaction between residues of the protein:
    - Bonded interaction
    - Non-bonded interaction

Understanding the relation between local geometrical properties of knotted proteins and their topology is a difficult task. It is a remarkable feature of the primary structures of proteins that establishes a set of native and non-native interactions to attain the correct topology in knotted proteins: i.e. specific short and long range interaction between the residues of polypeptide chains can encode the global informations of proteins such as tertiary structure and topology.

We use a structural-based coarse-grained model of proteins and introduce a protocol which by stochastic search in angular elastic parameter space, it provides

the essential set of bending and torsional elastic properties of the chain that encode the topology of the short knotted proteins.

In the aforementioned study in chapter 2, the knots were shallow and the most parts of the polypeptide chains formed the knot. However, in some knotted proteins with long polypeptide chain the topology is localized. Regarding to such a deep knots in proteins, intuitively, we suggest that non-bonded interactions in the long polypeptide chain may provide the knot in the early stage of folding, and thereafter, the proteins can fold into the native structure and make the knot localized. In chapter 3, we study the impact of two pairs of adhesive sites (two on the termini and two along the chain) on the knotting probability of a polymer that can also be modeled as DNA. Making use of extensive molecular dynamics simulations, we show that there is a specific position of the adhesive sites that maximizes the knotting probability of the polymer chain.

In the previous chapter we showed that knots are frequent in sufficiently long DNA strand. In chapter 4, we study the interaction between two knots on a polymer chain with fixed ends. The knots can pass through each other and form intertwined states where one knot is inside the other knot. We found that the intertwined state of two knots with opposite chirality is favored in comparison to the intertwined knots with same chirality.

Additionally, in chapter 5 we study the interaction between two trefoil knots tied on the same stretched polymer at different bending stiffness and we investigate the impact of the chain geometry that is enforced by topology.

The rationale behind the favored intertwined state of the knots with opposite chirality, can be understood by probing the effect of crossing pattern of the entwined knots. For this purpose, in chapter 6 we study braids of entwined DNA rings. We demonstrate that the crossing pattern of braid play a crucial role in its structural and dynamical properties. In particular we show that the braid with non-identical crossing pattern, that the latter is specified by comparing the partitioned inner and outer cross sections, enforces a positive and stronger correlation between the entangled rings.

Chapter 2

Folding of small knotted proteins: insights from a mean field coarse-grained model

This Chapter has been published as a research paper in The Journal of Chemical Physics. It is reprinted here with permission from the publisher.

Saeed Najafi and Raffaello Potestio
Folding of small knotted proteins: Insights from a mean field coarse-grained model
J. Chem. Phys 143, 243121 (2015)

Abstract

A small but relevant number of proteins whose native structure is known features nontrivial topology, i.e. they are knotted. Understanding the process of folding from a swollen unknotted state to the biologically-relevant native conformation is, for these proteins, particularly difficult, due to their rate-limiting topological entanglement. To shed some light into this conundrum we introduced a structure-based coarse-grained model of the protein, where the information about the folded conformation is encoded in bonded angular interactions only, which do not favor the formation of native contacts. A stochastic search scheme in parameter space

is employed to identify a set of interactions that maximizes the probability to attain the knotted state. The optimal knotting pathways of the two smallest known proteins, obtained through this approach, are consistent with the results derived by means of coarse-grained as well as full atomistic simulations.

2.1 Introduction

The three-dimensional organization of proteins, commonly described in terms of the geometrical arrangement of their secondary structure elements (the fold), can be characterized also by the topological state of their backbone. In fact, while the vast majority of known protein structures doesn't show complex self-entanglement, a relatively small but non-negligible number of folds features a knotted topology. To date, the number of knotted proteins in the Protein Data Bank [154] as recorded in the KnotProt database [72] amounts at circa 800, with varying degree of buriedness, localization and complexity. The latter, specifically, spans with decreasing population from the simplest knot, the 3_1, up to the most complex protein knot observed so far, a 6_1 [40].

Knotted structures were initially dismissed as undesired misfolded conformations or kinetic traps. The first systematic protein knot survey, carried out by Mansfield in 1994 [93], returned only one 'loosely formed' (shallow) knot, thereby raising perplexity about the existence of proteins with deeply entangled backbones. Subsequently, since the discovery of the first of such conformations [147], protein knots have been recognized as legitimate albeit surprising native folds. The rationale behind the self-entangled structure of these proteins has been related in some cases to an increased stability of the enzymatic active site [148] and in general of the native state [111], in other cases it has been interpreted as a harmless mutation [125]. No unified picture, however, has emerged to explain the presence of a knot in these biomolecules.

On the other hand, the predominance of unknotted native folds has been perceived as equally puzzling. In the previously mentioned work [93], Mansfield attributed the (quasi) absence of knotted proteins to a non-ergodic folding process. More recent investigations [85, 164], however, suggest that the ordered structure dictated by protein sequences tends to reduce the occurrence of knotted native states. In contrast, those proteins possessing a knot have, at least in some cases, de-

veloped specific sequences and structural elements tailored to fold in an entangled conformation [125, 60].

Besides of the biological role played by the knotted conformation of a protein, then, possibly even more complex to explain is the process that leads to it. This problem has been studied both experimentally [89, 91, 88, 83] and by means of computational approaches [62, 111, 161, 145, 125, 40, 166, 25, 47]. Experimental investigations carried out for the YibK and YbeA proteins [89, 90, 92, 83] have shown that their denatured state preserves the knotted topology, and that refolding from this state to the native conformation can occur efficiently. It was also shown that folding from a fully unknotted conformation is possible as well, albeit on larger time scales, thus pointing at the knotting process as the rate limiting step of the folding. This behavior complies with the mechanism, proposed by Mansfield [94], by which the knotted state is inherited from the denatured protein and preserved during the collapse towards the globular state. A crucial observation is that the knotting event is preceded by a well-defined, specific sequence of steps, as it was reported e.g. in an experimental study of Lim *et al.* [83] on YibK and YbeA. The same property, namely a single dominant folding/knotting mechanism, has been observed also in the context of all-atom Molecular Dynamics (MD) simulations of another protein, MJ0366 [25]; in contrast, simulations of coarse-grained (CG) models generally feature a higher plasticity of the folding landscape, that is, a variety of alternative folding pathways.

CG models have been often used to investigate the folding of knotted proteins [157, 161, 145, 117, 116, 46, 164]. The fewer degrees of freedom and the simpler, smoother interactions in fact allow a reduction of the typical simulation times required to observe the process. However, the success rate of these approaches in the context of knotted proteins is relatively limited [161, 40, 156]. Specifically in the case of models relying on the formation of the sole native contacts, such as Gō models [64], the reason of this fairly low native folding rate can be attributed to the frustration due to the high topological complexity of knotted proteins. In fact, the knotting process requires an ordered folding pathway, and most off-pathway arrangements represent kinetic traps. To obviate to this *faux pas* the chain has to backtrack [145, 117, 111, 60] to a more swollen state and try anew to fold along the correct path. Gō models are particularly sensitive to this property, because interactions built upon the proximity of residues in the native state are prone to establish prematurely the corresponding contacts, which then

have to be undone by backtracking. This phenomenon has been made particularly clear by the studies performed by Soler *et al.* [111], who demonstrated that the folding (and unfolding) efficiency of trefoil and pentaknot lattice proteins is greatly enhanced by structural mutations. The latter consist in the disruption of native contacts that occur on regions critical for knotting, e.g. the threading terminus and/or the knotting loop, yet are not directly involved in the folding process. By removing these interactions the untimely formation of contacts and the occurrence of kinetic traps is decreased, thereby reducing the need for backtracking.

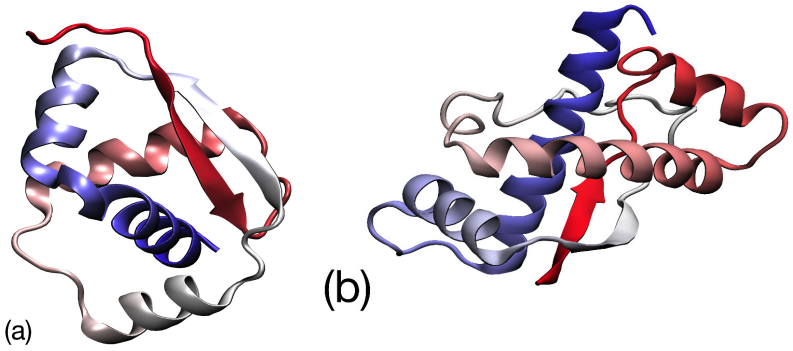


Fig. 2.1 Cartoon representation of the two knotted proteins studied in this work, namely the hypothetical protein MJ0366 (a), which is the smallest known protein knot, and the Agrobacterium Tumefaciens virC2 protein (b).

Furthermore, native contacts-based models exclude from the game non-native interactions, i.e. interactions between amino acid pairs that are not in contact in the native conformation. This approach is generally sound since, e.g. according to the results of Best and coworkers [36], these interactions do not play a major role in the folding process of various proteins. This is, however, in contrast with recent findings [61, 25, 47] indicating that non-native interactions are indeed a common trait in the folding process of knotted proteins. The reason for this discrepancy is likely to be found in the necessity, introduced by the nontrivial topology, to avoid *untimely* access to regions of the conformational space where the formation of native contacts, a rather positive step in an untangled protein, would represent a major mistake for a knotted protein.

Summarizing, the picture that is emerging out of the body of investigation carried out so far *in vitro* as well as *in silico* thus contains the following elements:

- **polarization:** the topological entanglement, i.e. the formation of the knot, occurs at a late stage of the folding process. However, it has to be preceded by a well-defined sequence of events, leading to a simple, polarized pathway.
- **non-native interactions:** these play a relevant role in the promotion of the knotted state, inasmuch as they can prevent the untimely formation of native contacts before the correct topology has been established.
- **optimality:** since the polarized, ordered pathway is prone to kinetic traps and backtracking, it also has to be optimized for folding efficiency.

Building on these observations, we explored the possibility of constructing a coarse-grained model of knotted proteins that aptly entails the characteristic features of these systems' folding process. Specifically, we aimed at devising a model of minimal complexity yet capable of providing useful information about the most likely knotting path followed by the protein. To this end, we considered a chain of beads, each representative of an amino acid, possessing only excluded volume and angular -i.e. bending and torsion- interactions with the neighbors in sequence. The latter potentials are employed as a basis set of functions on which the whole complexity of the intramolecular interactions of the real system is projected. The reference angles of these potentials are obtained from a target conformation, e.g. the PDB crystal structure or one close to the latter. This choice of interaction potentials is free from the bias towards the native contacts: in fact, this model is based on a particular, target conformation of the protein, yet the collapse towards it is only driven by concerted but *local* rearrangements of the chain.

Another pivotal premise is the optimization of the folding process. As previously observed, the folding pathway of a knotted protein is simple and characterized by a specific sequence of steps, likely resulting from the evolutionary pressure to maximize the folding probability of a structure with an extra degree of complexity with respect to unknotted proteins. Consequently, the interaction parameters of the model, i.e. the strengths of bending and torsion potentials, have been obtained by means of a stochastic search aimed at maximizing the successful knotting rate, a strategy that is reasonable to assume has guided also the evolutionary selection of these proteins.

This optimality criterion thus provides us with an unbiased force-field that, within the limitations of the specific interactions employed in the model, aims at providing a *mean field* description of the folding process of a knotted protein: along the route towards the target, entangled conformation, the out-of-path moves leading to kinetic traps are suppressed, and the knotting pathway is the most direct possible. The potential energy functions on which this model is based are not realistic representatives of the elaborate interactions to which a protein is subject, hence the dynamics it produces cannot be expected *a priori* to entail the full complexity of folding. Nonetheless, this model can yield valuable insight about the *kinematics* of the knotting process, that is, the optimal sequence of moves the chain has to perform to collapse into its reference, entangled conformation.

In the next section, the details of the method and the parametrization procedure are provided. The model, that we dub Elastic Folder Model (EFM), is first validated on an unknotted globular protein, namely Adenylate Kinase (PDB code 4AKE); subsequently, two cases are investigated and discussed: the hypothetical protein MJ0366 (PDB code 2EFV) and the Agrobacterium Tumefaciens virC2 protein (PDB code 2RH3), whose native structures are reported in Fig. 2.1.

2.2 Methods

The EFM employs a minimalistic representation of the protein based on the sole C_α atoms, connected to their first neighbors by means of stiff bonds. The only non-bonded interaction to which the centroids are subject is a short-ranged excluded volume, enforcing steric hindrance and preventing the chain from crossing itself. This tube-like model of the protein is then provided with bending and torsion potentials, whose reference angles are parametrized on a target structure. The latter could be, in principle, the PDB crystal structure; however, a slightly more swollen conformation has been employed in our study, as it will be discussed hereafter.

By construction, the reference structure represents the global minimum of the EFM; hence, the latter, being based on the folded (and knotted) conformation, falls in the category of structure-based models. It is important to stress, however, that the interactions in our system do not favor the formation of native contacts: the effective energy is minimized only by the onset of the target *local* arrangement of the residues.

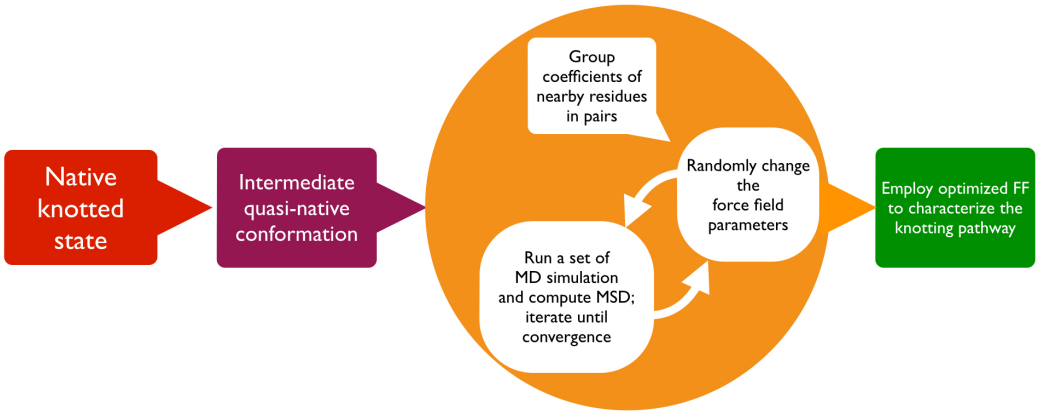


Fig. 2.2 Graphical representation of the iterative optimization procedure employed to parametrize the CG model. The choice of the elastic coefficients is performed solely on the basis of the successful collapse into the target conformation, so that no bias favoring native or non-native contacts is introduced in the model parametrization.

The reference angles represent the sole input parameters introduced in our model. The strength of the bending and torsion potentials are in fact determined by a Monte Carlo (MC) search, in which the set of parameters maximizing the successful collapse in the reference state is obtained. The details of the model and the MC optimization algorithm, graphically illustrated in Fig. 2.2, are described in the following subsections.

2.2.1 Coarse-grained model

Following a principle of extreme generality, the core of the EFM is a simple Kremer-Grest model [65]: the protein chain is described as a collection of identical beads, each representing an amino acid and centered on its C_α atom, connected by FENE bonds [65]. The only non-bonded interaction among the beads is a purely repulsive Weeks-Chandler-Anderson (WCA) [163] potential. Triplets and quadruplets of subsequent beads interact *via* bending and torsion potentials, respectively.

The total potential energy of the system is:

$$\mathscr{V} = U_{\text{WCA}} + U_{\text{FENE}} + U_{\text{bending}} + U_{\text{torsion}} \tag{2.1}$$

The WCA potential is given by:

$$\begin{aligned} U_{\text{WCA}} &= \frac{1}{2}\sum_{(i,j),j\neq i}^{N} V(d_{i,j}) \\ V(r) &= \begin{cases} 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + \frac{1}{4}\right] & \text{for } r \leq 2^{1/6}\sigma \\ 0 \text{ otherwise} \end{cases} \end{aligned} \tag{2.2}$$

and the FENE potential reads:

$$U_{\text{FENE}} = -\sum_{i=0}^{N-2} \frac{\kappa_{fene}}{2}\left(\frac{R_0}{\sigma}\right)^2 \ln\left[1-\left(\frac{d_{i,i+1}}{R_0}\right)^2\right] \tag{2.3}$$

where $d_{i,i+1} = |\vec{r}_i - \vec{r}_{i+1}|$ is the distance of the bead centers i and $i+1$, R_0 is the maximum bond length and κ_{fene} is the FENE interaction strength. Note that the diameter σ of the beads is taken as the length unit, and corresponds to the separation between two consecutive C_α atoms (roughly 3.8Å); all distances are expressed in reduced units. The WCA strength ε is taken as the energy unit, and the FENE bond constant κ_{fene} was set to 30ε as it is customary for the Kremer-Grest model [65].

The bending and torsion potentials are:

$$\begin{aligned} U_{\text{bending}} &= \sum_{i=1}^{N-2} \kappa_i^{bend}(\theta_i - \theta_i^0)^2 \\ U_{\text{torsion}} &= \sum_{i=1}^{N-3} U_i^{tor} \\ &\text{with:} \\ U_i^{tor} &= \kappa_i^{tor}\left(\cos(\phi_i - \phi_i^0) + \frac{1}{3}\cos(3(\phi_i - \phi_i^0))\right) \end{aligned} \tag{2.4}$$

θ_i^0 and ϕ_i^0 are the bending and torsion angle of the i-th bead in the reference states, respectively. κ_i^{bend} and κ_i^{tor} are the corresponding bending and torsion stiffness, respectively. Both types of coefficients were set, before the optimization, to 50ε, a sufficiently strong value to keep the angle fluctuations in the target conformation within 1 degree.

The potential of Eq. 6.2 is employed to perform overdamped Molecular Dynamics simulations in implicit solvent by means of the following Langevin equations of motion:

$$-\frac{\partial \mathscr{V}_i}{\partial \vec{r}_i(t)} - m\gamma\vec{v}_i(t) + \vec{R}(t) = 0 \tag{2.5}$$

where m, v_i, $\mathscr{V}_i$, γ, R_i and r_i are the mass, velocity, local potential energy, friction coefficient, random force and coordinate of the i-th bead, respectively. The equations of motion of the system are integrated with a symplectic, first order algorithm. The system parameters are listed in Table 2.1.

Parameter	Value
m	1
ε	1
MD time unit $\tau_{MD} = \sigma\sqrt{m/\varepsilon}$	1
Δt $[\tau_{MD}]$	$5 \cdot 10^{-4}$
R_0 $[\sigma]$	1.5
κ_{fene} $[\varepsilon]$	30
τ_{fricit} $[\tau_{MD}]$	1
$\kappa^{bend}, \kappa^{tor}[\varepsilon]$ (initial value)	50

Table 2.1 System parameters.

2.2.2 Quasi-native target structure

The EFM is characterized by a coarse description of the protein structure and very simple interactions, which are not sufficient to, and indeed not expected to, reproduce the whole complexity of the folding process: our aim is in fact to obtain the most efficient path connecting a completely unstructured conformation to the knotted state. To this end we employ as target structures, rather than the native PDB crystal structure, slightly swollen conformations geometrically and topologically close to the native state, that is, having a small root mean square distance (RMSD) from it and featuring the same knot. These *quasi-native structures* possess the same topological complexity of the native ones, but are less compact. This procedure allows us to interpret the bonded potential of the EFM as a mean-field approximation to the real intramolecular interactions that drive the folding until the knotted topology has established.

To obtain the quasi-native conformations we set up the CG model of Eq. 6.2 with angle parameters obtained from the PDB crystal structure and elastic parameters of equal strength for all angular potential energy terms; this model is then employed in a MD simulation at a temperature $T = 0.5\varepsilon$, which is verified *a posteriori* to be sufficient to induce a mild swelling of the structure, yet without changing the topological state with respect to the native conformation. For each knotted protein under examination we extract, from the pool of swollen structures, a subset of six conformations, each of which is employed as a target state for 1000 simulations at $T = 0.1\varepsilon$ with uniform elastic parameters starting from stretched, quasi-linear conformations. Out of the six target states, the *single one* that will be subsequently employed in our study is chosen as the one maximizing the probability to attain the target structure.

2.2.3 Monte Carlo search of the optimal model parameters

The following step consists in an iterative refinement of the set of elastic parameters $\{\kappa_i^{bend}, \kappa_i^{tor}\}$, to which we shall refer as the *force field* of the model. The refinement is performed through a Monte Carlo exploration of the parameter space aimed at maximizing the successful collapse into the target conformation.

The single Monte Carlo move is performed as follows. The value of the elastic coefficients is constrained in the range $[5\varepsilon, 500\varepsilon]$; at each step, the index and type (i.e. bending or torsion) of a coefficient are chosen at random, and the selected elastic parameter is added a value extracted from a uniform distribution in the range $[-2.5\varepsilon, 2.5\varepsilon]$. As the number of possible different force fields grows exponentially with the number of residues, we resorted to the following simplification in order to reduce the size of the combinatorial problem: instead of assigning to each residue independent bending and torsion coefficients, we constrained pairs of neighboring residues to possess identical values of the elastic parameters (each type separately). Each move then attempts to change the coefficients of the randomly chosen *pair* of residues.

During an optimization cycle, 8 independent MD simulations are concurrently ran; the chain attains either the target state or a local free energy minimum from which it cannot backtrack, thereby remaining "stuck". The simulations then stop when the RMSD with respect to the target state remains constant for a sufficiently

long time. Specifically, the RMSD is measured every $25\tau_{MD}$: if its values in two consecutive measures differ by less than 0.05σ the simulation is interrupted.

The average MSD between the configurations at the end of the simulations and the target state is then measured; the new set of parameters is accepted with probability $P = min\{1, \exp(-(\mathscr{F}_{new} - \mathscr{F}_{old}))\}$, with $\mathscr{F} = \text{MSD}(\text{final}|\text{target})/\sigma^2$.

2.2.4 Computation of the mean field free energy

In order to quantitatively characterize the pathway leading the protein chain towards the target structure, we have computed, for both MJ0366 and virC2, a *mean field free energy* landscape, reported and discussed in the following sections. This quantity is defined as:

$$\frac{F(x,y)}{k_B T} = -\ln P(x,y) \tag{2.6}$$

$P(x,y)$ is the probability distribution of conformations in a reduced space of collective coordinates. These conformations are obtained, for each protein, from 1000 trajectories, irrespective whether successful or not. The simulations were ran at a temperature $k_B T = 0.1\varepsilon$, a value at which the knotted proteins in the target conformation stably fluctuate about the reference structure with a RMSD $\leq 1\sigma$. The functional form of the collective coordinates (x,y) is system-specific, in general however they correspond to the RMSD between the instantaneous and the target conformations of the whole protein chain or part of it.

2.2.5 Knot analysis

The analysis of the topological state of the protein chains is performed by means of the algorithm implemented in the `KNOTFIND` [68] package. To circularize the protein chain we made use of the minimally interfering closure algorithm introduced by Tubiana *et al.* [152].

2.3 Results and discussion

2.3.1 Validation: unknotted Adenylate Kinase protein

To validate the EFM, and in particular its ability to collapse an unfolded chain towards the desired target structure, we carried out simulations of the single chain, open conformer of the Adenylate Kinase enzyme (PDB code 4AKE). This is a 214 residue, globular protein responsible for the energetic balance of the cell; the absence of topological entanglement in the fold and its relatively small size -yet more than double with respect to the knotted proteins under examination in the following- make it a good case to put our model at test. We employed a setup in which all elastic coefficients were given the same values, namely $\kappa_i^{bend} = \kappa_i^{tor} = 50\varepsilon$, and the temperature was set to $T = 0.2\varepsilon$. The PDB crystal structure has been employed as a target state. Each simulation was started from a stretched, linear configuration.

As many as 70% of the 200 independent simulations collapsed into the target state. In Fig. 2.3a we report the RMSD between the instantaneous configuration and the native structure as a function of time for a typical successful trajectory. The polymer smoothly collapses towards the target conformation without particular features: only at $t \sim 400\tau_{MD}$ the RMSD shows a very short, almost negligible plateau. The normalized total energy time series, inset of Fig. 2.3a, highlights brief, intermediate states of the collapse, of which the 'jump' at $\sim 600\tau_{MD}$ is the last; these sudden jumps in the energy can be associated to small rearrangements of the local structure.

These results indicate that this very simple model is capable of driving a relatively large protein towards its target structure. The absence of any topological complexity results in a straightforward collapse with no intermediate transient states.

2.3.2 Case study 1: MJ0366 knotted protein

The first nontrivial case considered is protein MJ0366 (PDB code 2EFV, Fig. 2.1a), which, with only 82 amino acids and a shallow, right-handed trefoil knot distributed along the whole length, is the shortest known knotted protein [40]. Because of

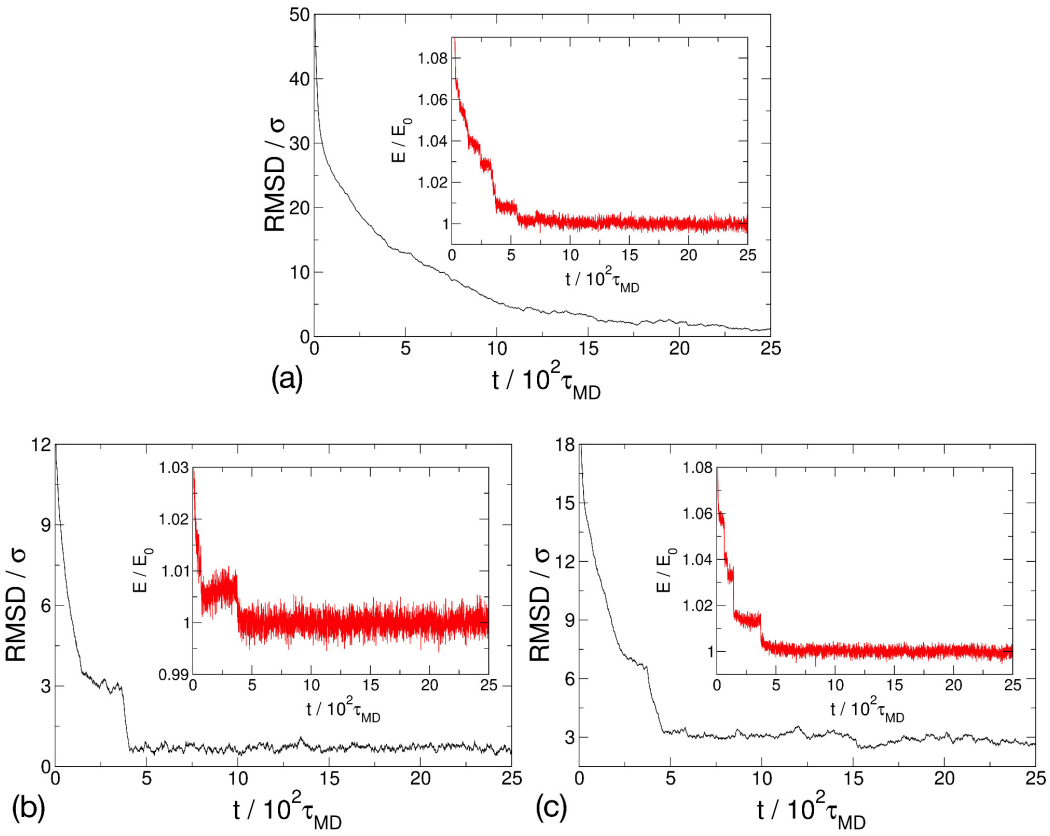


Fig. 2.3 RMSD with respect to the native state (main plot) and normalized total energy (inset) as a function of time during a typical collapse trajectory of Adenylate Kinase (a), MJ0366 (b), and virC2 (c).

its small size, this polypeptide is well-suited to computational investigation, and indeed has been the object of a number of coarse-grained as well as full atomistic studies.

As a first step we obtained a knotted target conformation following the protocol described in the Materials and Methods section. This structure is slightly more swollen than the PDB crystal structure, having a RMSD of only 1σ with respect to it, yet preserves all its fundamental geometrical as well as topological features. The target state was employed to perform 1000 independent simulations with a uniform-parameters force field. The success rate of this model amounted to 45%.

In order to increase this already fairly high figure, we refined the force field by means of the stochastic optimization procedure previously discussed. The improved quality of the resulting set of elastic coefficients, whose values are plotted in Fig. 2.4a and provided in Table S1 of the Supplemental Material, was

confirmed by the higher fraction of successful simulations, that was as large as 89%.

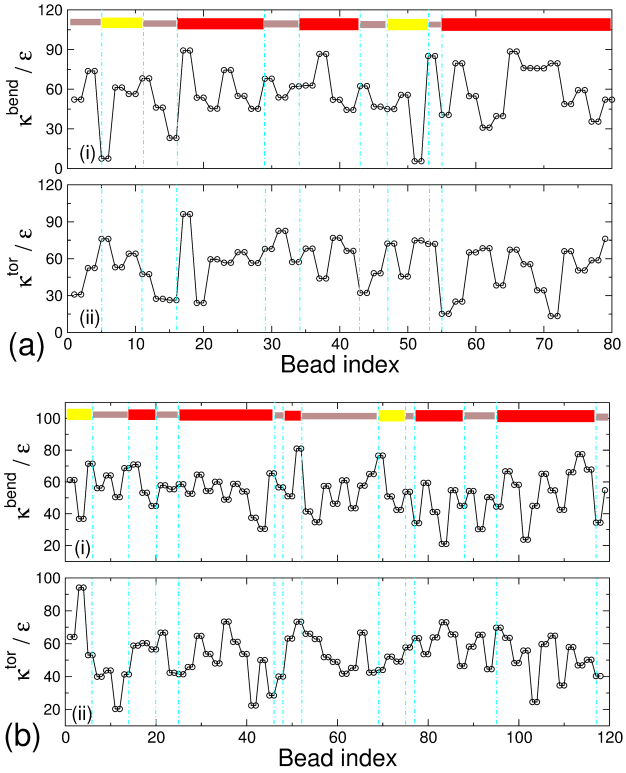


Fig. 2.4 Elastic coefficients of MJ0366 (a), and virC2 (b). Each panel shows the value of the optimal elastic parameter for bending (subpanels *i*) and torsion (subpanels *ii*) as a function of the bead index. Above each plot the secondary structure elements of each protein are indicated, according to the following color code: red for the α helices, yellow for the β strands, and brown for the unstructured regions. The light cyan dashed lines are a guide to the eye.

In Fig. 2.3b we report the RMSD as a function of time for one of the runs correctly attaining the target conformation. At variance with respect to the case of 4AKE, here we can distinctly observe a transition in the RMSD, which clearly correlates with a jump in the total energy. This sudden drop in the RMSD, preceded by a short yet noticeable interval in which the decrease proceeded at a slower pace, is suggestive of the existence of an intermediate state. In Fig. 2.5 (Multimedia view) we illustrate a few relevant snapshots of a knotting trajectory (the full video is provided as integral multimedia). The representation highlights the conformational transitions the protein undergoes to reach the knotted target state: in each subpanel (a-j) the structure before and after the transition is shown, to emphasize the specific

changes that occurred. Additionally, the final conformation of each step is colored according to the average strength of the elastic coefficients of the optimized force field shown in Fig. 2.4a; the colors go from red through white to blue following increasing values.

As a first event we observe a rapid, global formation of the secondary structure elements, especially the helices. Subsequently (panels b-c) two regions with high values of the elastic parameters, specifically residues $17-18$ and, on a lesser note, residues $65-66$, form sharply bent turns. The collapse proceeds with the stabilization of the "white" regions having intermediate elastic strength (panel d); further on, the weaker linker following the second α helix (residues $43-45$) reaches its target conformation (panels e-f), thereby closing the hinge and compacting the structure. Finally, the segments having weakest elastic parameters (red) close down the terminal helix into the knotting loop, establishing the entangled topology.

As expected, the time sequence of the events that lead the chain to relax into the target conformation is aptly encoded in the strength of the angular potential's elastic coefficients, and the time order correlates with it: regions required to attain earlier their target structure possess stronger elastic parameters. The optimization procedure modulates the coefficients in order to impose the most efficient sequence of events to the relaxation process.

Previous work on MJ0366 [25, 47] has demonstrated that its folding is indeed characterized by a well-defined sequence of steps, and the folding mechanism follows a consistent, polarized path. Specifically, by means of atomistic simulations it was observed that the creation of a loop, stabilized by the formation of a native β-sheet, invariably precedes the onset of the knotted state by threading of the protein C terminus through the aforementioned loop. A plot of the β-sheet RMSD *vs.* the total RMSD for the trajectories successfully folding towards the native state showed [25] that the closure of the loop occurs when the global RMSD is still as large as 12Å, and the knot formed in most cases after the loop.

To quantitatively compare this behavior to that of our coarse-grained system, we performed a similar analysis on the trajectories obtained with the EFM. In subpanel 1 of Fig. 2.7a we report the mean field free energy (see Methods section) projected on the two collective variables represented by the total RMSD and the RMSD relative to the β-sheet. These results are consistent with those obtained

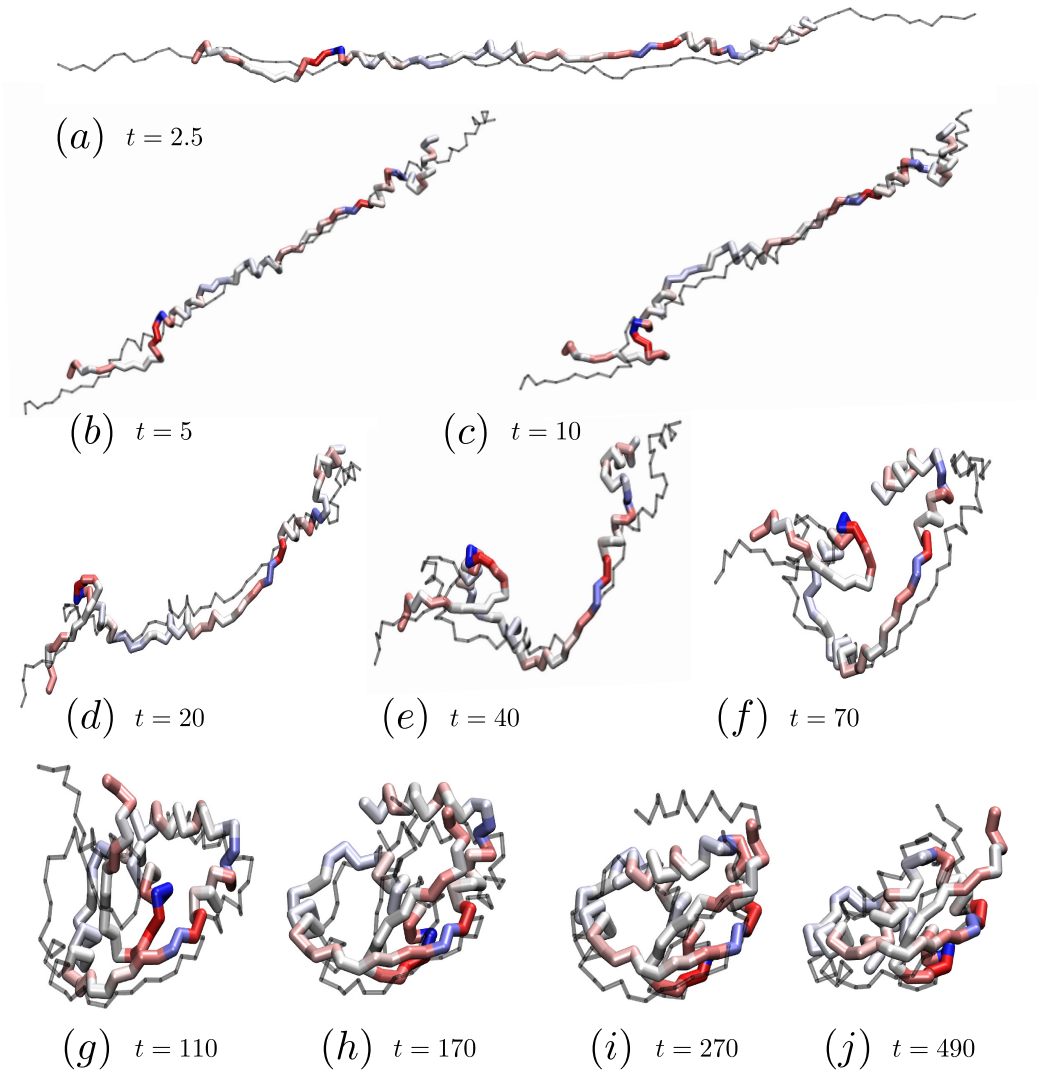


Fig. 2.5 Optimal knotting process of protein MJ0366 as obtained from a simulation employing the refined force field of the EFM. The full video is available as integral multimedia (Multimedia view). Each snapshot (a-j) shows a relevant conformational transition from a given conformation to another. At each step, the thin gray trace represents the structure before the transition, while the thicker one is the structure after the transition. The final conformation of a given step is the initial one of the following, and the time (in units of τ_{MD}) provided at each transition indicates the onset of the final conformations. The initial structure in panel (a) corresponds to $t = 0$. The color coding on the final structures maps the local strength of the elastic parameters, measured as the average of the bending and torsion coefficients of each bead. The color coding goes, for increasing values, from red through white to blue. See main text for the discussion of the knotting pathway.

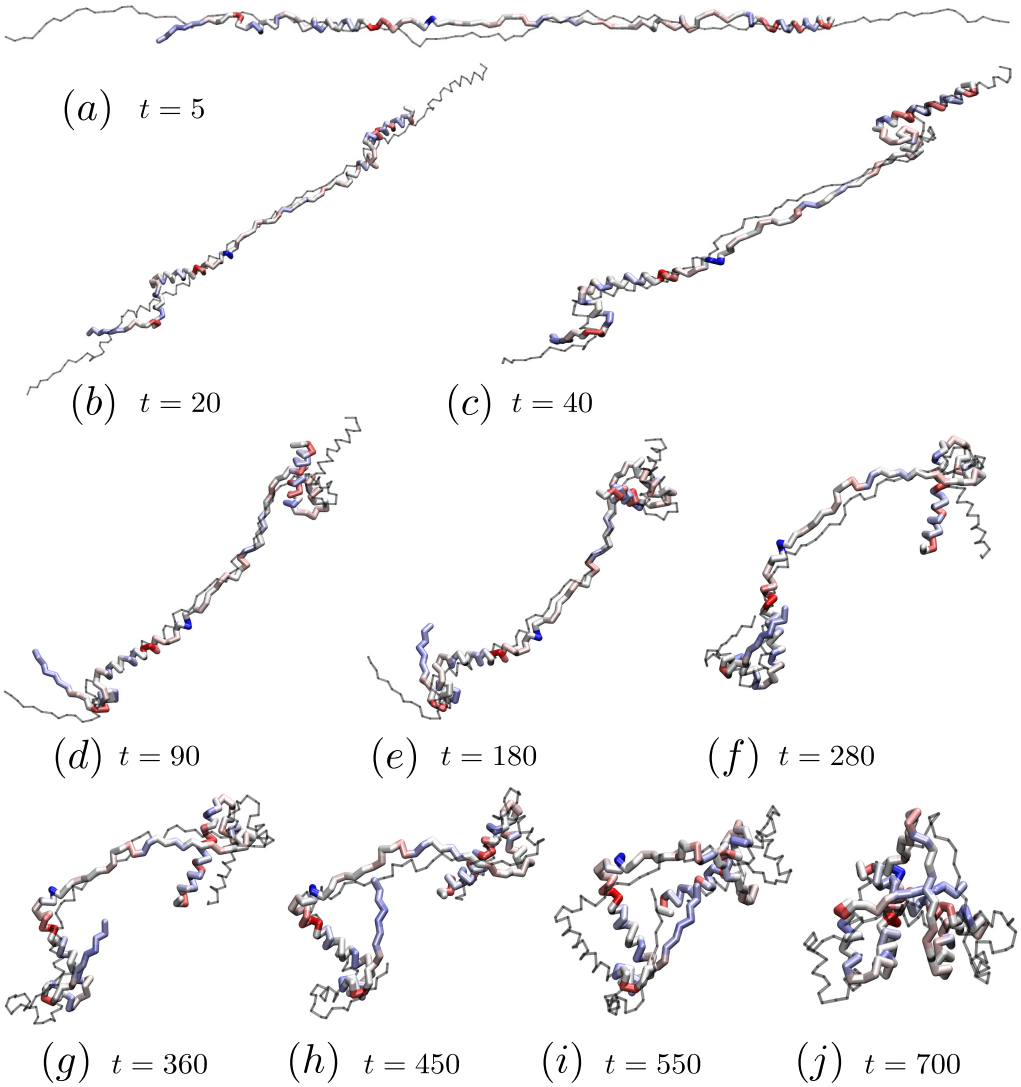


Fig. 2.6 Optimal knotting process of protein virC2 as obtained from a simulation employing the refined force field of the EFM. The full video is available as integral multimedia (Multimedia view). As in Fig. 2.5, the relevant conformational transitions occurring during the knotting are illustrated. The color coding on the final structures maps the average elastic parameters, with increasing strength going from red through white to blue.

by means of full atomistic simulations. In fact, we observe a flat path leading to a shallow local minimum, which corresponds to the formation of the β-sheet loop ($\sim 1\sigma$) while the global RMSD is still relatively large ($\sim 3\sigma$). From these minima the system can overcome a barrier of $\sim 2.5 k_B T$ and proceed to a second basin related to the final stages of the collapse; during this process, the knot is finally formed by threading the C-terminal helix through the loop closed by the β-sheet.

The absence, in the EFM, of any native contact interaction, and the smooth relaxation dynamics towards the target conformation determine a less sharp separation between the time scales of the two collective variables with respect to the atomistic simulations. It is however remarkable that the crude, simplified model here discussed is capable of reproducing the overall behavior observed in simulations employing a much more accurate and detailed description of the system.

We can thus draw two main conclusions from these results. First, the capability to systematically and reproducibly form a molecular knot can be embedded in local angular interactions that do not account for the global geometry/topology of the system. Second, we observe that, by construction, the folding pathway of MJ0366 is the most efficient one (in terms of knotting probability) given the underlying model: in fact, this pathway was determined by a force field obtained by means of a stochastic optimization criterion, and it is consistent with the one produced by atomistic simulations. Hence, the folding process of this protein likely evolved to follow a one-way-track pathway towards the native conformation.

It is worth noting here that, as already observed by other authors, the folding pathway of MJ0366 is characterized by the presence of an intermediate state: this fact is quite unusual for proteins this small size, typically featuring a two-state folding mechanism. This property appears as a common feature of knotted proteins [161, 89, 25, 83], and can be rationalized in terms of the necessity to construct a folding pathway entailing a unique, well defined sequence of steps. In this conditions backtracking is reduced, yet at the expenses of a broader variety of paths connecting unfolded and folded states.

We conclude the discussion about the MJ0366 protein with a reference to the specific mechanism it employs to form the knot. In fact, in the all-atom simulations performed by a Beccara *et al.* [25] three possible ways in which the knot can establish have been observed: the dominant *threading* mechanism, in

which the C-terminal α helix pierces unbent through the loop; the *slipknotting* [117] mechanism, where the tip of the helix is bent backwards; and the *mouse-trap* mechanism, in which the loop snaps back onto the already formed helix. On the contrary, simulations employing C_α Gō models [145] or all-atom resolution Gō models [117] seem to favor the slipknotting process.

In the simulations performed with the EFM we uniquely observed the threading mechanism: this fact can be understood considering that the target structure contains a well-formed C-terminal helix. The latter forms very early (see panels a-c of Fig. 2.5), since the terminal residues are more free to move with respect to the ones in the center of the chain, and they can quickly reach the reference conformation. We care to remark, however, that the threading mechanism emerges, in the case of the EFM, from a combination of model-specific and target structure-specific features, and should not be taken as an evidence to proof or confute one knotting mechanism over another.

2.3.3 Case study 2: virC2 knotted protein

The second case under study is the C-terminal domain of Agrobacterium Tumefaciens virC2 protein (PDB code 2RH3, Fig. 2.1b). The overall fold of this 121 residue long trefoil protein is similar to that of MJ0366, featuring a loop 'sealed' by a β-sheet, a strand of which is located close to the N terminus. A helix, constituting the C terminus, pierces through the loop thereby forming the knot.

The structural as well as topological similarity of these two proteins is suggestive of comparable folding mechanisms. Indeed, simulations [143] employing a coarse-grained, native contacts-based model of virC2 have highlighted a similarity of its free energy landscape with the corresponding one of MJ0366. In these simulations, the free energy was monitored in terms of two collective variables, namely the overall folding progress (analogous to the total RMSD with respect to the native state) and the folding progress of the sole β-sheet. The landscapes of both proteins featured a prolate well favoring the formation of the sheet and, in proximity of conformations corresponding to a stably formed loop, a deeper minimum centered onto the native state. However, the way these two basins are connected is different between the two proteins. In fact, the free energy landscape of protein MJ0366 imposes virtually no barrier in going from an unfolded conformation where the

β-sheet has formed towards the native state basin; on the other hand, the sheet-forming basin of virC2 is less pronounced and, most notably, separated from the native state minimum by a $\sim 5\ k_BT$ high barrier. Given this pattern of similarities and differences, it is legitimate to hypothesize that the knotting mechanism of MJ0366, namely the early closure of the loop followed by the α-helix threading, represents *one possible* folding pathway for virC2 protein, yet maybe not the only one nor necessarily the most favorable.

Some insight in the understanding of virC2 folding can be provided by the optimal pathway obtained by means of the EFM. The coarse-graining procedure applied to virC2 returned a force field, plotted in Fig. 2.4b and provided in Table S2 of the Supplemental Material, successfully driving the protein to the knotted target conformation, which has a RMSD $< 3\sigma$ with respect to the native structure, in 82% of the cases. The total RMSD and total energy of the system *vs.* time during a successful knotting trajectory, reported in Fig. 2.3c, highlights, also in this case, the presence of an intermediate state. In Fig. 2.7b we illustrate the relevant milestones of this trajectory, and in Fig. 2.6 (Multimedia view) a more detailed time series of the knotting event can be appreciated; the full video is available as integral multimedia. Visual inspection of these data shows that the collapse process followed by this protein is not similar to that of MJ0366: in fact, the β-sheet closes at a very late stage, when the knot has already formed. Hence, the β-sheet RMSD is not, in this case, the optimal collective variable to pair to the global RMSD.

To better understand this process in terms of the collective variables appropriate to it, we consider the RMSD of a chain subset that, as observed in the trajectory, reaches very early a stable conformation, that is, the stretch of the chain roughly corresponding to the C-terminal half. This segment quickly forms a loosely open loop, not coinciding with the one bracketed by the β strands. A plot of the mean field free energy in terms of total and loop RMSD, as reported in subpanel 1 of Fig. 2.7b, shows that the collapse rapidly proceeds towards the stabilization of the C-terminal part of the loop, including the helix; after a short transient, during which the total RMSD decreases at roughly constant loop RMSD, the remaining segment of the loop snaps back in a mouse-trap like mechanism (analogous to the one observed in few atomistic simulation trajectories of MJ0366 [25, 47]) by which the β-sheet loop and the knot are concurrently formed.

The pathway followed by virC2 modeled with the EFM suggests an alternative mechanism that can be employed by the protein to reach its topologically entangled native state. The assumptions embedded in the model, namely the maximal efficiency in collapsing the protein chain into the target state and the optimality criteria employed to parametrize the force field, provide the basis to expect this pathway to be the most efficient and straightforward. Most notably, this expectation is consistent with the results obtained by means of more conventional coarse-grained simulations [143], which suggest that the β-sheet closure of the loop as in MJ0366 is made difficult by the presence of a non negligible free energy barrier; the latter, on the contrary, is absent in the case of MJ0366, for which our model indeed finds a knotting pathway analogous to the one observed in atomistic simulations.

2.4 Conclusions

Understanding the folding mechanism of knotted proteins is a major endeavor. Two reasons constitute the most part of the difficulty to reach this objective, a practical one and a conceptual one. The practical reason is the fact that the knot formation represents a rate-limiting step, which determines, for entangled proteins, longer folding time scales with respect to similarly sized but unknotted ones. In order to follow the process by means of simulations, larger computational resources are hence required, dramatically limiting the access to realistic trajectories for molecules larger than a few tens of amino acids.

The second, more profound limitation is represented by the conceptual difficulty to understand how a system driven by relatively simple interactions can perform a task that demands to a human a nontrivial degree of planning and accurate execution. An unknotted folding pathway might be hard to dissect and describe, but it is not difficult to ultimately interpret it as a downfall along the free energy landscape. The folding of a knotted protein is somewhat closer to a mountain-to-valley excursion with well planned stop-overs through necessary intermediate steps.

In the present work we developed a very crude coarse-grained model aimed at providing relevant information about the knotted protein folding with minimal computational expenditure. Our model relies on two assumptions: that the complex interactions driving the process can be projected onto, and approximated with,

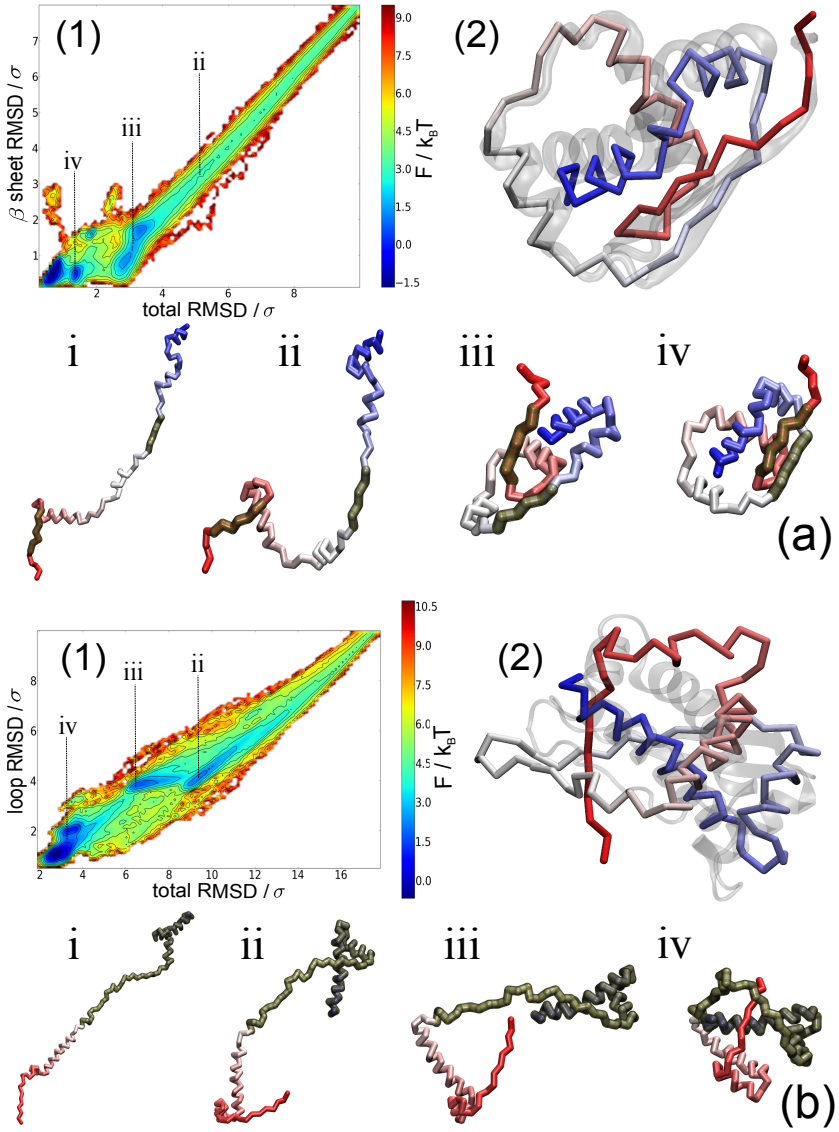


Fig. 2.7 Analysis of the collapse pathway for MJ0366 (a) and virC2 (b) towards their respective target conformations. Each panel shows the mean field free energy landscape (subpanels 1) as measured in terms of the collective variables best suited to describe their folding pathways (see text for a detailed description). The target conformation of each protein is shown in subpanels 2 in tube representation; the corresponding native state aligned onto this structure is shown in transparent gray cartoon representation. A few intermediate steps of the process are illustrated by the corresponding snapshots of collapse trajectories of MJ0366 (as in Fig. 2.5 (Multimedia view)) and virC2 (as in Fig. 2.6 (Multimedia view)) provided as integral multimedia. The collective coordinates corresponding to the most relevant 'milestones' are indicated in the free energy plots with letter labels referring to the trajectory snapshots. The chain segments employed to define the second collective coordinate (β-sheet RMSD for MJ0366 and loop RMSD for virC2) are highlighted in dark green transparent material.

effective potentials local in space and sequence; and that the free energy funnel leading towards the native state is straightforward and polarized. The specific parameters of the force field were obtained by means of an unsupervised, stochastic optimization maximizing the probability to attain the target conformation; yet, the potential energy of our model is not build to favor the formation of native contacts. Hence, the forces driving the proteins toward their knotted conformations are a mean field proxy of both native as well as non-native interactions.

This approach was proven to be capable of reproducing the optimal folding pathway followed by the shortest known knotted protein, MJ0366, consistent with realistic all atom simulations. In the next-to-simplest case, represented by the virC2 protein, the folding mechanism returned by the model here discussed is different from what is observed in native contacts-based, coarse-grained simulations; however, these studies have also shown the presence of a free energy barrier absent in the case of MJ0366, thus suggesting that the folding mechanism of the latter might be accessible to virC2, yet subleading with respect to a more effective pathway.

In conclusion, it is reasonable to assume that the folding strategies developed by knotted proteins are highly optimized in order to reduce the impact of the rate-limiting topological complexity. The knowledge accumulated by studies employing coarse-grained as well as more realistic computational descriptions have unveiled the delicate interplay of native and non-native interactions and the polarization of the folding pathways. In the future, further insight might be provided by simplified models focusing on the optimality of the knotting mechanism, that is emerging to be the unifying trait of knotted proteins.

2.5 Acknowledgments

The authors are deeply indebted with M. Sandal, T. Bereau and K. Kremer for an attentive and critical reading of the manuscript, and wish to thank A.C. Fogarty for introducing them to the art of Python scripting for plots. The Rechenzentrum Garching is gratefully acknowledged for the computer simulation resources.

Chapter 3

Two adhesive sites can enhance the knotting probability of DNA

This Chapter has been published as a research paper in PLoS ONE journal. It is reprinted here with permission from the publisher.

Saeed Najafi and Raffaello Potestio
Two Adhesive Sites Can Enhance the Knotting Probability of DNA
PLoS ONE 10, e0132132 (2015)

Abstract

Self-entanglement, or knotting, is entropically favored in long polymers. Relatively short polymers such as proteins can knot as well, but in this case the entanglement is mainly driven by fine-tuned, sequence-specific interactions. The relation between the sequence of a long polymer and its topological state is here investigated by means of a coarse-grained model of DNA. We demonstrate that the introduction of two adhesive regions along the sequence of a self-avoiding chain substantially increases the probability of forming a knot.

3.1 Introduction

In the last few decades, concepts from Topology have increasingly gained ground in the study of biopolymers, most notably in the case of proteins [147, 157, 91, 25, 40, 125] and DNA [123, 28, 27, 159, 107, 99, 127, 109]. These molecules, in fact, can undergo the same fate of an everyday piece of rope: they can be knotted. Characterized by a wealth of three-dimensional conformations and traditionally described in structural terms, biopolymers have demonstrated to entail a similarly rich variety of topological features, which largely affect their behavior [31, 124]. Knotted protein folds, for example, have long been associated only with folding errors [93, 75], whether *in vivo* or *in silico*. To the date of writing, the KnotProt [72] database lists about 800 protein entries with a proper knot (i.e. no slipknots), the functional relevance of which is often under debate. On the contrary, a long polymer chain such as DNA, with a much smaller degree of sequence-dependence with respect to a protein, is expected to be knotted with high probability [52, 160, 144]. It is hence not surprising to find experimental as well as numerical evidence of topological entanglement in DNA strands confined in viral capsids [27, 107]. On the other hand, it comes to much surprise that the 100 Mbp-long genetic material contained in a human chromosome is knot-free [130]. Recently, the same puzzling absence of knots has been observed also in RNA molecules [104]. The well established biology paradigm *sequence* $\rightarrow$ *structure* $\rightarrow$ *function* [44] is then enriched with topology, and it is of the greatest importance to understand the interplay between these four instances. To shed light on this conundrum, a body of work has been carried out especially by means of numerical simulations, with focus on different aspects of the problem, for example knotted protein folding [161, 146, 78, 40, 25], DNA [159, 130, 107, 99, 127], knot-specific sequences in model polymers [45, 46, 164], and equilibrium properties of knotted chains [160, 144, 110, 106, 120, 124, 31, 151].

In the present article, we focus our investigation on the relation between sequence and topology of DNA. Genetic material, in fact, is at the center of a broad range of topology-related biological and material science problems. The self-recognition of complementary sequences allows the formation of nontrivial two- and three-dimensional structures. This property lies at the core of genetic recombination, and enables the occurrence of secondary structure element formation in RNA macromolecules, as is in the case e.g. of ribosomes. The possibility to form

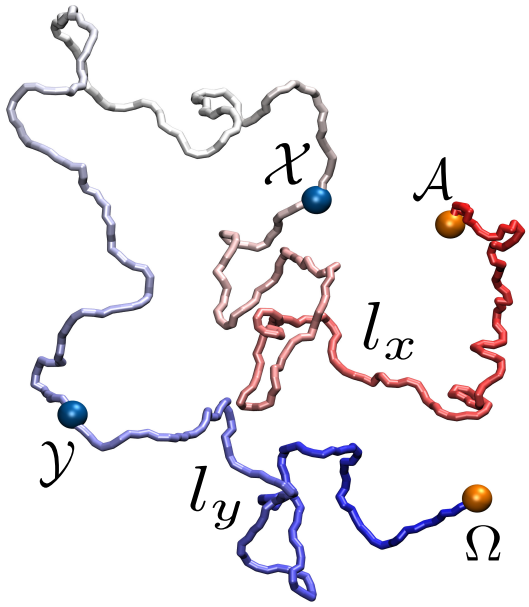


Fig. 3.1 **Snapshot of the DNA model under exam.** The orange beads, labeled $\mathscr{A}$ and Ω, indicate the sticky patches at the termini of the polymer, while the blue beads, labeled $\mathscr{X}$ and $\mathscr{Y}$, represent those along the polymer chain. These two pairs of regions do not attract each other. The chain segment comprised between $\mathscr{A}$ and $\mathscr{X}$ (resp. $\mathscr{Y}$ and Ω) has length l_x (resp. l_y).

selective, sequence-dependent bonds has been widely exploited in the field of structural DNA nanotechnology to (self-)assemble DNA origami [139, 140, 42, 136], DNA superlattices [86], Borromean rings [96] and complex-shaped nanoscale objects [132], such as gears, stars and smileys. At a higher level of genetic material organization, we find that the three-dimensional architecture of the 30 nm chromatin fiber is largely affected by the formation of loops [69, 38]. These structures are stabilized by protein complexes, e.g. the CTCF transcription factor, selectively bridging specific binding sites along the DNA sequence [77, 134, 67]. The formation of these loops plays a crucial role in the regulation of gene expression.

Natural knots in the genetic material remain nonetheless elusive. As previously noted, in fact, a survey of RNA molecules indicated a remarkable absence of knotted structures [104], in spite of the capability, in principle, to exploit secondary structure formation to achieve topologically entangled folds akin to those observed in knotted proteins. Similarly, in the case of nuclear chromatin the length of the fiber (which largely exceeds that of the CTCF-induced loops), the activity

of topology-regulating enzymes such as topoisomerases [84, 43, 138], and non-equilibrium dynamics [130] make it impossible for these loops to become elements of topological entanglement, i.e. to knot the chromosome. However, the mechanisms underlying loop formation in both aforementioned cases are completely general, and, if properly designed and applied to shorter fibers, could be exploited to manipulate their knotted state.

Here we investigate what impact the formation of a loop can have on the topology of a DNA fiber. Specifically, we consider a filament of double-stranded DNA (dsDNA) modeled as a chain of beads with excluded volume [123, 130, 99, 82], and introduce two pairs of adhesive monomers, as illustrated in Fig. 6.1, which permanently stick to each other when sufficiently close. One pair of such monomers (labeled $\mathscr{A},\Omega$) is located at the termini, and has the role of circularizing the polymer to freeze its topological state. A second pair of adhesive monomers (labeled $\mathscr{X},\mathscr{Y}$), not interacting with the first, is located along the chain. The latter is initially set up in an open, linear conformation. A constant-temperature molecular dynamics (MD) simulation is carried out until the termini become close enough to stick and cyclize the polymer. The knotted state of the resulting conformation is then analyzed as a function of the position of the $\mathscr{X},\mathscr{Y}$ sticky regions. Two types of DNA models, termed L-DNA and S-DNA, are employed, both composed by polymer chains having the same number of monomers but different persistence length. The first case models a dsDNA chain long enough so that its persistence length is negligible; the second case corresponds, for dsDNA at physiological salt conditions (0.15 M NaCl) [135], to a 7.5 kbp long filament, roughly the length of the papillomavirus genome [50]. Further details on the model and the simulation protocol are reported in the Material and Methods section.

3.2 Results and discussion

The sticky monomers $\mathscr{X},\mathscr{Y}$, located along the chain, identify two types of sub-chains: the *arms* of the polymer, namely the chain segments between a terminus and the closest sticky bead along the sequence ($[\mathscr{A}-\mathscr{X}]$ and $[\Omega-\mathscr{Y}]$); and the *sticky loop*, i.e. the segment comprised between the central sticky beads ($[\mathscr{X}-\mathscr{Y}]$). The chains are composed by $N=500$ beads, along which the sticky monomers can be located only at distances $l_x = 50n_x,\ l_y = 500-50n_y$ from the termini,

with $n_x, n_y = 1, 2, \cdots 9$. Of the $9 \times 9 = 81$ possible configurations we have to exclude the ones in which the beads coincide and, for symmetry, the pairs in which $l_x = l'_y$, $l_y = l'_x$. This leaves us with 20 non-redundant locations of the sticky bead pairs along the chain. In Fig. 3.2 the configurations are distributed in the l_x, l_y plane; along the positive-tangent diagonals (parallel to the `A-B` line) one has configurations in which the length of the arms changes but it is the same for the two arms, while along the negative-tangent diagonals (parallel to the `C-B` line) the size of the loop stays constant, but its location along the chain changes. An illustration of the three extreme cases is also provided.

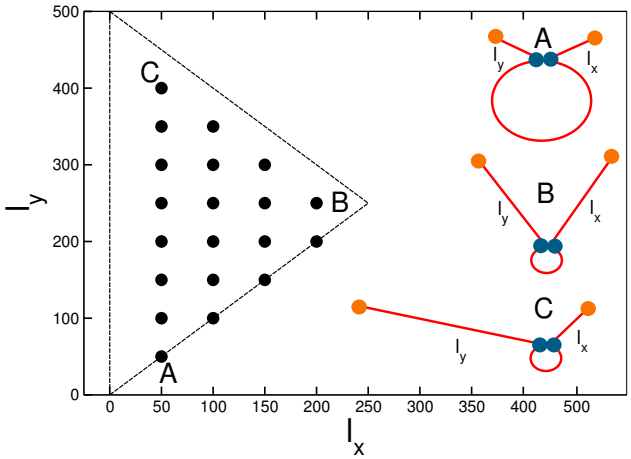


Fig. 3.2 **The parameter space explored in the present study.** The coordinate on each axis represents the length of an *arm*, i.e. the number of polymer beads between a terminus and the closest sticky bead along the sequence. The dots indicate the values that have been investigated. In the right half of the plot, three sketches of the polymer with closed sticky loop illustrate the `A`, `B`, `C` points.

For each of the 20 locations of the adhesive monomers we measured the relative knotting probability (RKP) of the polymer, defined as:

$$\mathrm{RKP} = \frac{P_K(l_x, l_y)}{P_K^0} \tag{3.1}$$

where $P_K(l_x, l_y)$ is the knotting probability of the chain with both types of sticky monomers, and P_K^0 is the (reference) knotting probability of an equivalent chain with adhesive termini only. These probabilities are computed as the fraction of knotted final configurations over the total. For the two types of chain under exam, namely the L-DNA and the S-DNA, we performed $104 \cdot 10^3$ and $24 \cdot 10^3$ independent simulations, respectively; as discussed in the Materials and Methods

section, the larger number of runs for the L-DNA with respect to the S-DNA is required by the smaller knotting probability of the former over the latter [158].

As anticipated in the Introduction, the final configuration is defined as the last frame of a MD simulation; the latter is interrupted when the terminal sticky beads cyclize the chain, irrespective of whether the $\mathscr{X}, \mathscr{Y}$ sticky monomers are joint or not. It has to be stressed that the presence of the adhesive termini restricts the conformational space the polymers can sample. The circularization of the chain, in fact, freezes the topology and prevents the polymer from changing its knotted state, either by tying a more complex knot or by untying the existing one. This constraint, however, reduces the complexity of the phenomenology under exam and removes any possible source of ambiguity due to the detection of the knotted state of an open chain [152].

The scheme in Fig. 3.2 allows us to predict the qualitative behavior of the RKP as a function of the sticky monomer locations. First, we expect to observe an overall increase of the knotting probability with respect to a chain without central adhesive monomers: these, in fact, favor more dense configurations, thus "self-confining" the polymer and enhancing its propensity to entangle [105]. Second, in the configurations corresponding to the corners of the triangle the sticky beads are very close to the termini (point `A` in Fig. 3.2) or to each other (point `B`) or both (point `C`). Therefore, their effect on the topology will be negligible, and the RKP should be almost unity in those points. Along the `A-B` and `A-C` segments the size and position of the sticky loop can vary. For Rolle's theorem [128], then, we expect the presence, along these sides of the triangle, of a local maximum of the probability. The case for the `C-B` side is different: along the latter, in fact, the location of the sticky beads changes, but the size of the loop remains the same - zero in the limiting case in which the two sticky monomers coincide. Hence, the RKP should remain fixed at unity.

The predicted behavior is confirmed by the heat-map plots shown in Fig. 3.3a-b, which report the RKP for L-DNA and S-DNA, respectively. As expected, for all points we have RKP$\geq$ 1, which indicates an overall enhancement of the knotting propensity in presence of the sticky monomers. The maximum relative increase amounts to 11.6 for the L-DNA and 3.9 for the S-DNA.

A remarkable result is that, in spite of a noticeable difference in the absolute numbers, the qualitative behaviors of the RKPs of L-DNA and S-DNA are decidedly

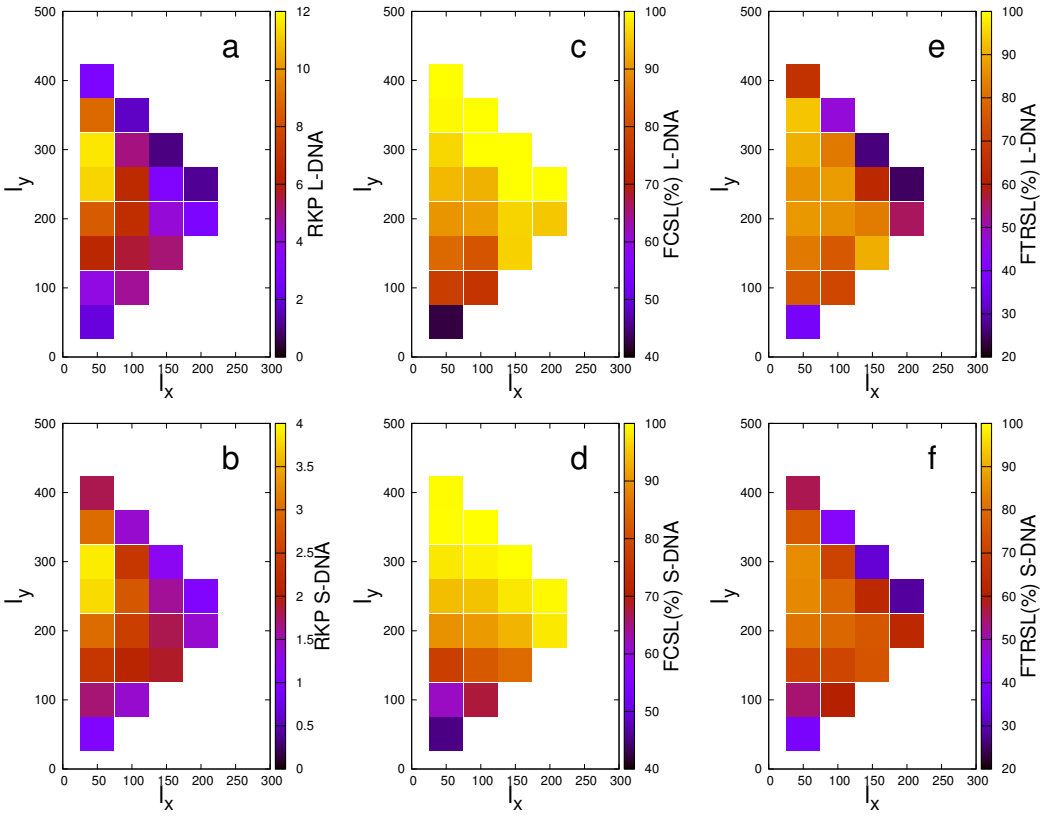


Fig. 3.3 **Heat-map representation of the simulation data as a function of the sticky monomer location.** Top row: data for the L-DNA. Bottom row: data for the S-DNA. RKP: relative knotting probability. FCSL: fraction of knotted chains with closed sticky loop. FTRSL: fraction of knotted chains with topologically relevant sticky loops.

similar. Not only, in fact, both probability distributions comply with the expected presence of maxima along the `A-B` and `A-C` segments; additionally, the absolute maximum of the RKP occurs in the same point of the plane, namely the $\{50, 300\}$ point, for both chain types. The fact that this particular pair of arm lengths maximizes the knotting probability results from the interplay of two competing effects: on the one hand, the size of the loop has to be sufficiently large in order to allow and favor the threading of the arms through it; on the other hand, an adequate length difference between the arms has to be guaranteed, so that one of them is short enough to pierce through the loop before the terminal sticky monomers come close together and freeze the topology in an unknotted state. As a matter of fact, in this configuration the $\mathscr{X}, \mathscr{Y}$ sticky beads are only 150 monomers apart, so that they can "find" each other before the termini do the same. At the same time, the loop is large enough to allow either arm to go though it (if it were perfectly circular,

its diameter would be roughly 50 beads). Consistently with these observation, we also note that the relative maximum of the RKP along the A-B line (that is, for equal length of the arms) is located at the $\{150, 150\}$ point: in this configuration, the sticky loop is 200 beads long, very close to the optimal length.

As already noted, the location of the maxima is the same in the two cases under exam, irrespectively of the persistence length of the chain. If we assume the same fine-grained system to underlie both models, we can rephrase this observation by saying that DNA chains of different length feature the same optimal location of the sticky patches. From this perspective, we thus understand the optimization of the knotting probability by means of the introduction of adhesive regions as a property that only depends on their position relative to the chain length, with the latter only affecting the absolute value of the enhancement of the knotting probability.

Further insight comes from the analysis of the knot complexity in the two cases under exam (numerical values are reported in S1 Table). In fact, the vast majority ($> 98\%$ on average) of the knots observed in the L-DNA are 3_1, while in the S-DNA case the fraction of knots more complex than a 3_1 can be larger than 8.4%. A possible explanation for this tendency towards complex knots in the S-DNA is that some degree of bending rigidity allows for larger, more open loops that favor multiple threading. More detail is provided in Fig. 3.4, where we report the knot spectrum for the two chain types cumulated over all possible sticky monomer location. (The break-down of the spectrum for all 20 sticky monomer locations under exam is provided in Fig. B in S1 Text) From the S-DNA data we see that the 5_2 knot type is more abundant than the 5_1 by approximately a factor 2. The same trend is observed [107] in simulations of DNA under confinement. The knotting mechanism in the aforementioned condition cannot, obviously, be compared with the knotting experienced by the chains under exam in the present work. It is however interesting to observe how the same behavior can emerge by means of the *self-confinement* introduced by the formation of the sticky loop. In the L-DNA case we observe the opposite balance between 5_1 and 5_2 knots, but their abundance is too small to rule out an insufficient sampling.

The data reported in Fig. 3.3a,b show that, in general, the presence of the sticky monomers along the chain increases the relative knotting probability with respect to equivalent chains having only the sticky termini. This observation alone, however, does not provide any information about the role played by the formation

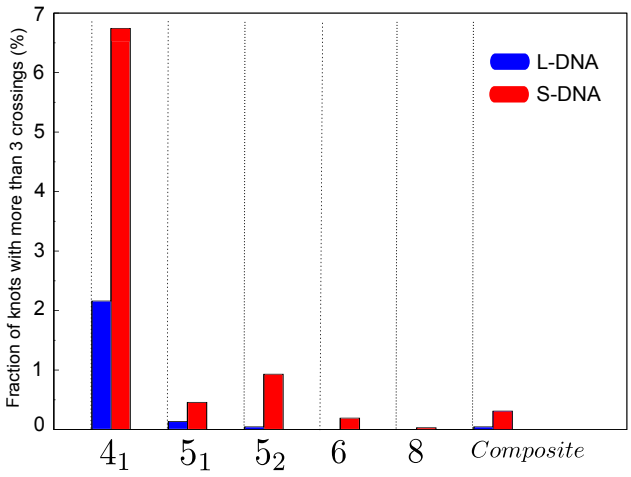


Fig. 3.4 **Cumulative knot spectrum of the L-DNA (blue) and S-DNA (red) chains, irrespective of the sticky monomer location.** The percentage refers to the total number of knotted configurations. The fraction of 3_1 knots is $\sim 98\%$ and $\sim 91.5\%$ for L-DNA and S-DNA, respectively.

of a sticky loop. To ascertain this, we measured how many knotted configurations involve the latter. These values are graphically illustrated in Fig. 3.3c,d for the L-DNA and S-DNA, respectively. The fraction of knotted chains in which the loop is closed ranges between 84% and 100% for 31 of the 40 cases under exam, depending on the sticky monomer location. More specifically, and not surprisingly, in both L-DNA and S-DNA we observe a positive gradient in the fraction of closed loops in the direction of shorter distances between the $\mathscr{X}$, $\mathscr{Y}$ beads. In other words, the closer the sticky beads are along the sequence, the higher the probability that they will adhere. For the sticky bead location maximizing the RKP, namely the $\{50, 300\}$ point, the fraction of knotted configurations involving a closed sticky loop amounts to 96.6% for L-DNA and 98.1% for S-DNA.

Having assessed that the sticky loops are present in most of knotted configurations, we need to discriminate between the case in which the loop and the knot coexist without interfering and the case in which loop and knot are topologically entangled. These two possibilities are depicted in Fig. 3.5. Given a closed, knotted chain featuring a sticky loop, we deem the latter to be *topologically relevant* if its removal determines a change in the topology of the chain (panels a to c). Alternatively, we consider the sticky loop irrelevant to the formation of the knot if the latter is completely localized within the sticky loop or in the complementary loop

obtained by removing the former (as illustrated in panels d to f). A more detailed discussion of this algorithm is provided in section C in S1 Text.

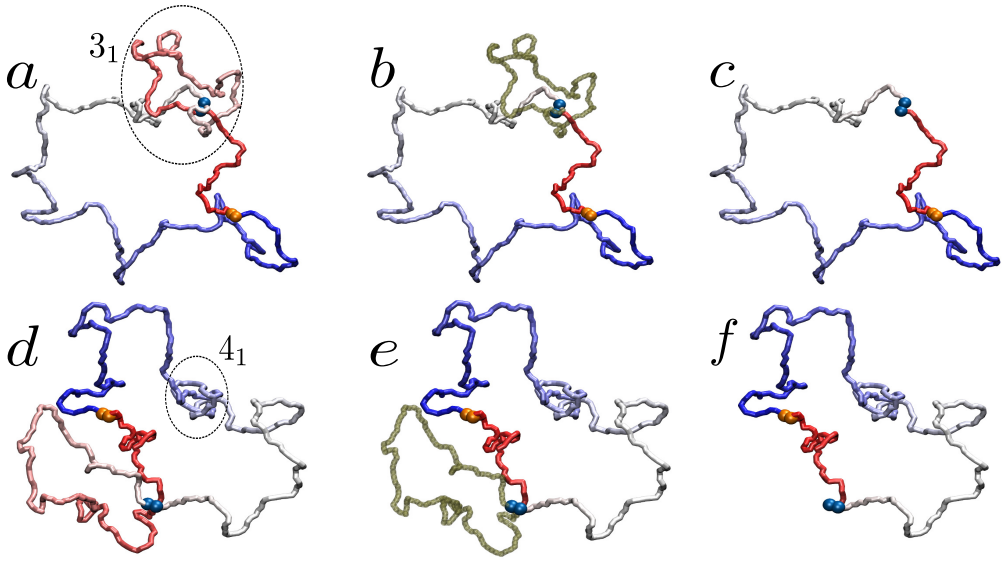


Fig. 3.5 **Illustration of the algorithm employed to determine whether a sticky loop is relevant or not for the topological state of the polymer.** If the sticky loop (dark green segment) is entangled with the knot (panels a and b), its removal determines a change in the chain topology (panel c). On the contrary, if no chain segment pierces the sticky loop (panels d and e), its excision does not modify the topological state of the whole polymer (panel f). In the figure, the sticky beads at the termini are orange, while those along the sequence are blue. In the first panel of both cases (a and d) the knotted region is inscribed in a circle. The two configurations are obtained from simulations of S-DNA with sticky monomers located at the $\{50, 300\}$ point.

As shown in Fig. 3.3e,f, the amount of loop-dependent knots is in general large among the different sticky beads locations, half of them having a percentage equal to or larger than 75%. Also much smaller values are present in some cases, though never below 25%. The highest values are registered in proximity of the RKP $\{50, 300\}$ maximum, namely 90.7% for L-DNA and 85.4% for S-DNA, thus suggesting that the formation of a stable loop in the appropriate region of the chain indeed enhances the probability to form a knot by threading a polymer terminus through it.

3.3 Conclusions

In summary, we have demonstrated that the knotting probability of a filament of dsDNA, modeled as a thick self-avoiding chain, can be increased by introducing, along its sequence, two adhesive regions. The extent of this enhancement depends only on the location of the latter, and its qualitative pattern is the same for the two chain types considered, namely with zero and finite persistence length.

The highest chance of self-entangling the polymer is obtained when the location of the adhesive regions optimizes the interplay of two competing effects: one is the enhancement of the threading probability, which is proportional to the loop size; the other is the polymer circularization event, which is more probable when the length difference between the two arms is small. Our data indicate that the knotting probability is maximized, in both L-DNA and S-DNA, when the sticky monomers allow the formation of a loop of length $\sim 1/3$ of the whole polymer located close to one of the termini.

Because of the generality of the model employed in this study, the presented results are prone to verification and employment on a fairly broad range of length scales: in fact, as mentioned in the Introduction, genetic material at different length scales -from single-stranded RNA to 30 nm chromatin fibers- is provided with the elements necessary to form loops. The knowledge of the mechanisms favoring the realization of knotted topologies can, therefore, be employed not only to design self-knotting structures, but also to rationalize the absence of entanglement in otherwise knot-prone systems.

Understanding how the sequence of a polymer determines not only its geometrical structure but also its knotted state is a relevant and difficult task. The model discussed here aimed at reducing the complexity of the problem to the core by introducing the smallest amount of sequence information in a plain polymer with excluded volume. The rich behavior featured by this simple system proved useful to build a basic understanding of the relation between the constituents of a complex molecule and its topological state, and provided the instrument for future work that will elucidate the knotting process.

3.4 Materials and Methods

3.4.1 Model and simulation details

For our study we employed the well-established Kremer-Grest model of a coarse-grained polymer [65]. Specifically, our polymer chain is modeled as a collection of identical beads of unit mass, connected by anharmonic FENE bonds [65]. The non-bonded interaction acting among the beads is a Weeks-Chandler-Anderson [163] (WCA) potential, which enforces the excluded volume. The only beads featuring a further non-bonded potential are the termini and the two sticky beads along the chain. The most general form of the total potential energy of the chain is:

$$\mathscr{H} = U_{WCA} + U_{FENE} + U_{stick} + U_{bend} \tag{3.2}$$

The WCA potential is given by:

$$U_{WCA} = \frac{1}{2} \sum_{(i,j),j\neq i}^{N} V(d_{i,j}), \tag{3.3}$$

$$V(r) = \begin{cases} 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + \frac{1}{4} \right] & \text{for } r \leq 2^{1/6}\sigma \\ 0 \quad \text{otherwise} \end{cases} \tag{3.4}$$

where $\varepsilon = 1$ sets the energy scale. The FENE potential reads:

$$U_{FENE} = -\sum_{i=1}^{N-1} \frac{\kappa_{fene}}{2} \left(\frac{R_0}{\sigma}\right)^2 \ln \left[1 - \left(\frac{d_{i,i+1}}{R_0} \right)^2 \right] \tag{3.5}$$

where $d_{i,i+1} = |\vec{r}_i - \vec{r}_{i+1}|$ is the distance of the bead centers i and $i+1$, $R_0 = 1.5\sigma$ is the maximum bond length and $\kappa_{fene} = 30\varepsilon$ is the interaction strength.

The sticky interaction is modeled as a negative Gaussian short range potential acting between sticky beads of like type. The potential is given by:

$$U_{stick} = G(|\vec{r}_\Omega - \vec{r}_{\mathscr{A}}|) + G(|\vec{r}_{\mathscr{Y}} - \vec{r}_{\mathscr{X}}|) \tag{3.6}$$

$$G(r) = \begin{cases} -U_0 \exp\left(-\frac{(r-2^{1/6}\sigma)^2}{2\lambda^2} \right) & \text{for } 2^{1/6}\sigma \leq r \leq (2^{1/6}+5)\sigma \\ 0 \quad \text{otherwise} \end{cases}$$

where $\vec{r}_{\mathscr{A}}$, $\vec{r}_{\Omega}$, $\vec{r}_{\mathscr{X}}$ and $\vec{r}_{\mathscr{Y}}$ are the coordinates of sticky beads $\mathscr{A}$, Ω, $\mathscr{X}$ and $\mathscr{Y}$, respectively. The parameters $U_0 = 100\varepsilon$ and $\lambda = 2.5\sigma$ are chosen so that the interaction between sticky monomers is sufficiently strong, i.e. larger than thermal fluctuations, when they are closer than $2-3\sigma$.

The bending rigidity potential is defined as:

$$U_{bend} = \sum_{i=1}^{N-2} \frac{\kappa_{bend}}{2} (\theta_i - \pi)^2 \tag{3.7}$$

where θ_i is the angle formed by a triplet of consecutive beads with the *ith* bead at the center. The bending stiffness is $\kappa_{bend} = 10 k_B T$, with k_B Boltzmann constant.

The constant-temperature MD simulations are carried out with an in-house code integrating the Langevin equations of motion with $k_B T = \varepsilon$ and $\tau = \sigma\sqrt{m/\varepsilon} = 1$ MD time units. In the L-DNA case (resp. the S-DNA case), a total number of $104 \cdot 10^3$ (resp. $24 \cdot 10^3$) simulations have been performed for each of the 20 central sticky monomer locations. The factor ~ 4 separating the numbers of individual runs of the two sets depends on the different knotting probability between L-DNA and S-DNA, respectively $1.9417 \cdot 10^{-4}$ and $6.6680 \cdot 10^{-3}$. Since the latter is much higher than the former (as discussed e.g. in [158]) we performed a substantially larger number of runs for the L-DNA case in order to obtain a statistically significant number of knotted conformations for both polymer chain types.

3.4.2 Knot analysis

Knots are mathematically well-defined only for closed curves. As our simulation protocol automatically returns circularized conformations, we are spared from the need to perform a chain closure [152], a time-consuming and potentially ambiguous procedure (especially in the case of buried termini). The topological state of our closed chains has been obtained applying the algorithm implemented in the `KNOTFIND` package.

3.5 Supporting Information

3.6 Data summary

The numerical values of the results of the present work are listed in Table 3.1 hereafter.

3.7 Computation of the persistence length

The system under study is a 500 bead long polymer. The diameter σ of the beads is taken as a length unit, and the same distance separates the centers of consecutive beads. Because of the short-range WCA repulsion, the chain cannot cross itself. Two types of polymer are employed in this work, namely one in which no bending rigidity is present (the L-DNA), and one in which a three-body bonded potential imposes a given stiffness. These two types of chain thus feature different persistence length l_P, that is, the characteristic decay length of the bond vector self correlation.

In the first case the expression for l_P is given by:

$$l_P = -\frac{\sigma}{\ln(1-\frac{\sigma^2}{4\Delta^2})} \tag{3.8}$$

where σ is the bond length and Δ is the cross-sectional radius of the chain. In the case under exam, we have $\Delta \simeq \sigma/2$: with this value in Eq. 3.8 we obtain $l_P = 0$.

For the case in which a bending rigidity κ_{bend} is present (S-DNA), we impose its strength based on the requirement to reproduce DNA in physiological salt conditions, i.e. 0.15 M NaCl. We have that in this case the effective diameter of DNA is $\sigma = 5nm$. Assuming a persistence length of 50 nm or 150 bp we obtain the appropriate bending energy for this condition as:

$$\frac{\kappa_{bend}}{k_B T} \simeq \frac{l_P}{\sigma} = 10 \tag{3.9}$$

For a polymer of length $L = 500\sigma$ as in our case, we obtain $L/l_P = 50$, corresponding to $150bp \times L/l_P = 7500bp$.

3.8 Simulation protocol

In order to investigate how the presence and location of the sticky beads along the polymer chain affect the polymer's topology we performed different sets of simulations. Each of these corresponds to a particular location of the central sticky beads. Specifically, we placed the sticky monomers in all possible, non-redundant locations separated by 50 beads available on a chain of polymerization degree $N = 500$. This choice leaves us with 20 possible locations.

For each position of the central sticky monomers, two different blocks of simulations are performed, one for L-DNA and one for S-DNA. A single simulation runs according to the following protocol:

1. the polymer chain is initialized in a randomized, extended conformation
2. the Langevin equations of motion are integrated
3. as soon as the two terminal beads "stick together" through the Gaussian potential the simulation is interrupted - whether the central sticky beads have adhered to each other or not
4. the final configuration of the polymer chain is stored; the trajectory is discarded

3.9 Knot analysis

The topological state of our closed chains has been obtained applying the KNOTFIND algorithm [1].

In our analysis, we had to take into account the fact that the overall topological state of the chain might be independent of the loop formed by the central sticky monomers. In order to ascertain the entanglement between knot and sticky loop in a knotted configuration, we considered all those chains that have a nontrivial topology *and* a sticky loop, i.e., the central sticky beads $\{\mathscr{X},\mathscr{Y}\}$ have adhered before the termini did. We then identify all monomers in the sticky loop (that is,

[1] J. Hoste and M. Thistlethwaite, KNOTFIND, 1999, www.math.utk.edu/morwen/knotscape.html

all monomers having index $i \in (\mathscr{X}, \mathscr{Y})$) and the complementary loop formed by the two arms, and analyze their topology. If one of them is in the same knotted state of the full chain, we deduce that the knot has formed deep into the loop but not because of it, or equivalently on the arms. On the other hand, if both the sticky loop and the loop formed by the two arms alone have a different topology with respect to the full chain, the loop is deemed to be topologically relevant.

It is worth stressing here the the sticky loop can result to be topologically relevant whether it was necessary to the formation of the knot or not. The closure of the loop, in fact, can occur *after* the knotting of the chain and entangle it *a posteriori*, so that the loop removal would result in a crossing, e.g. as illustrated in Fig. 3.6. Our analysis, which is based on the sole final configuration, does not explicitly identify these cases, rather it provides a measure of the overall topological and geometrical entanglement in the polymer.

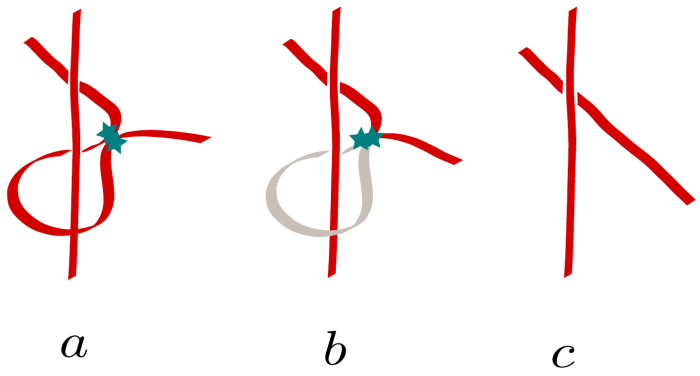


Fig. 3.6 Illustration of the process by which a the excision of a sticky loop results in the swapping of two chain segments. For this to happen, the sticky loop has to be pierced by another stretch of the chain.

3.10 Knot spectrum for all sticky bead positions

We report in Fig. 3.7 hereafter the break-down of the knot spectrum for the two DNA types, separated by sticky bead position. The $0-0$ label indicates the reference chains without central $(\mathscr{X}, \mathscr{Y})$ sticky beads.

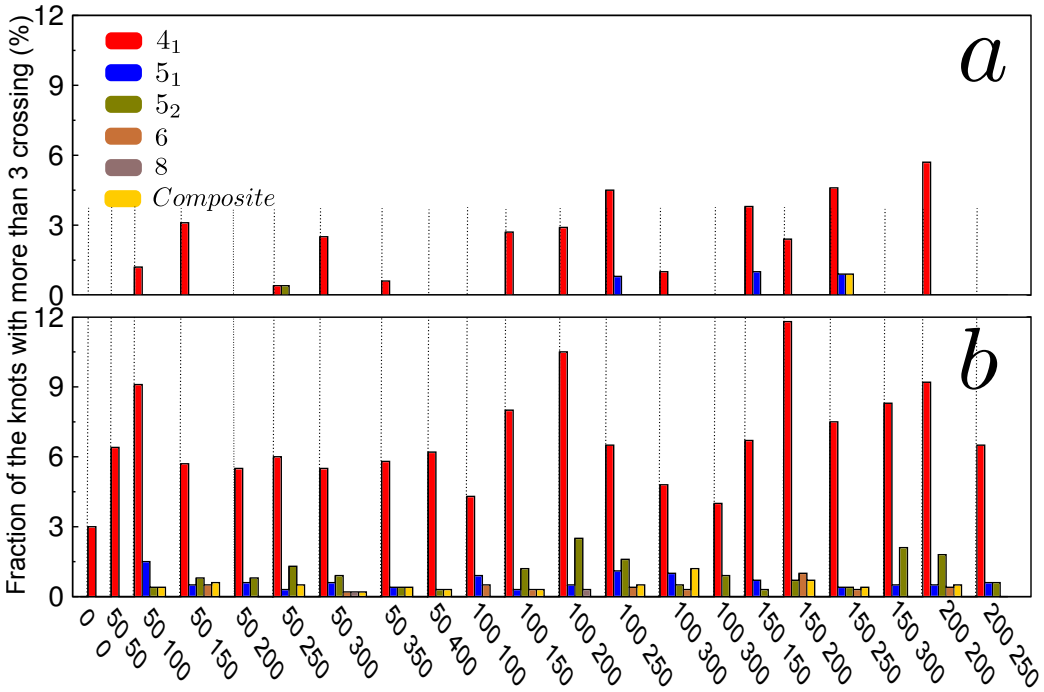


Fig. 3.7 Knot spectrum of L-DNA (panel a) and S-DNA (panel b). The $0-0$ label indicates the reference simulations where only the terminal sticky monomers are present.

3.11 Acknowledgments

The authors are deeply indebted with L. Tubiana for inspiring discussion and for providing the executable of the KNOTFIND package. We thank K. Kremer, K. Prakash and D. Donadio for a careful reading of the paper and valuable discussion. The Rechenzentrum Garching is gratefully acknowledged for the computer simulation resources.

		RKP		Knots with more than 3 crossings (%)		FCSL (%)		FTRSL (%)	
N	$l_x\ l_y$	**L-DNA**	**S-DNA**	**L-DNA**	**S-DNA**	**L-DNA**	**S-DNA**	**L-DNA**	**S-DNA**
1	50 50	2.0	1.0	—	4_1 (6.4)	42.5	45.2	37.5	38.2
2	50 100	3.9	1.7	4_1 (1.2)	4_1 (9.1), 5_1 (1.5), 5_2 (0.4), $3_1\#3_1$ (0.4)	77.2	61.7	75.9	54.2
3	50 150	6.4	2.4	4_1 (3.1)	4_1 (5.7), 5_1 (0.5), 5_2 (0.8), 6_1 (0.5), $3_1\#3_1$ (0.6)	84.5	77.6	82.2	71.7
4	50 200	8.5	3.0	—	4_1 (5.5), 5_1 (0.6), 5_2 (0.8)	90.2	89.8	87.3	81.6
5	50 250	11.3	3.8	4_1 (0.4), 5_2 (0.4)	4_1 (6.0), 5_1 (0.3), 5_2 (1.3), $3_1\#3_1$ (0.5)	93.9	94.3	86.5	85.1
6	50 300	11.6	3.9	4_1 (2.5)	4_1 (5.5), 5_1 (0.6), 5_2 (0.9), $3_1\#3_1$ (0.2), 6_3 (0.2), 8_{21} (0.2)	96.6	98.1	90.7	85.4
7	50 350	8.9	3.0	4_1 (0.6)	4_1 (5.8), 5_1 (0.4), 5_2 (0.4), $3_1\#4_1$ (0.4)	99.4	99.8	93.4	76.2
8	50 400	2.9	1.8	—	4_1 (6.2), 5_2 (0.3), $3_1\#3_1$ (0.3)	100	99.7	66.1	55.9
9	100 100	4.8	1.4	—	4_1 (4.3), 5_1 (0.9), 6_1 (0.5)	75.3	67.8	72.1	60.4
10	100 150	5.6	2.1	4_1 (2.7)	4_1 (8.0), 5_1 (0.3), 5_2 (1.2), $3_1\#3_1$ (0.3), 6_3 (0.3)	81.4	82.2	76.1	71.8
11	100 200	6.8	2.5	4_1 (2.9)	4_1 (10.5), 5_1 (0.5), 5_2 (2.5), 8_1 (0.3)	91.4	90.7	86.3	79.2
12	100 250	6.6	2.8	4_1 (4.5), 5_1 (0.8)	4_1 (6.5), 5_1 (1.1), 5_2 (1.6), 6_1 (0.2), $3_1\#3_1$ (0.5), 6_2 (0.2)	93.2	94.9	88.0	79.0
13	100 300	5.0	2.4	4_1 (1.0)	4_1 (4.8), 5_1 (1.0), 5_2 (0.5), $3_1\#3_1$ (0.8), $4_1\#4_1$ (0.3), 6_2 (0.3)	100	99.0	82.2	71.3
14	100 350	1.7	1.4	—	4_1 (4.0), 5_2 (0.9)	100	100	47.1	42.2
15	150 150	5.1	1.9	4_1 (3.8), 5_1 (1.0)	4_1 (6.7), 5_1 (0.7), 5_2 (0.3)	96.2	84.9	90.4	75.3
16	150 200	4.1	1.8	4_1 (2.4)	4_1 (11.8), 5_2 (0.7), 6_1 (0.7), $3_1\#3_1$ (0.7), 6_3 (0.3)	96.3	93.4	82.9	75.8
17	150 250	3.0	1.6	4_1 (4.6), 5_1 (0.9), $3_1\#5_2$ (0.9)	4_1 (7.5), 5_1 (0.4), 5_2 (0.4), 6_1 (0.3), $3_1\#3_1$ (0.4)	100	98.1	63.3	63.7
18	150 300	1.0	1.2	—	4_1 (8.3), 5_1 (0.5), 5_2 (2.1)	100	100	26.5	32.6
19	200 200	2.9	1.4	4_1 (5.7)	4_1 (9.2), 5_1 (0.5), 5_2 (1.8), 6_1 (0.4), $3_1\#3_1$ (0.5)	95.2	98.2	55.2	63.6
20	200 250	1.2	1.0	—	4_1 (6.5), 5_1 (0.6), 5_2 (0.6)	100	99.4	25.0	28.6

Table 3.1 Summary of the data presented in the present work. RKP: relative knotting probability. FCSL: fraction of knotted chains with closed sticky loop. FTRSL: fraction of knotted chains with topologically relevant sticky loops.

Chapter 4

Chirality modifies the interaction between knots

This Chapter has been published as a research paper in letter format in Europhysics Letters. It is reprinted here with permission from the publisher.

Saeed Najafi, Luca Tubiana, Rudolf Podgornik, and Raffaello Potestio
Chirality modifies the interaction between knots
EPL (Europhysics Letters) 114, 50007 (2016)

Abstract

In this study we consider an idealization of a typical optical tweezers experiment involving a semiflexible double-knotted polymer, with steric hindrance and persistence length matching those of dsDNA in high salt concentration, under strong stretching. Using exhaustive Molecular Dynamics simulations we show that not only does a double-knotted dsDNA filament under tension possess a free energy minimum when the two knots are intertwined, but also that the depth of this minimum depends on the relative chirality of the two knots. We rationalize this dependence of the effective interaction on the chirality in terms of a competition between chain entropy and bending energy.

4.1 Letter

The study of physical knots in polymers is an important emerging topic in biophysics and soft matter in general. Since the original conjecture that knots should be ubiquitous in sufficiently long chains [52, 63], later proved by Sumners and Whittington [144], knots have been found or tied in a variety of biopolymers, from DNA [28, 27, 99, 98] to proteins [147, 157, 79, 125, 166, 25, 114] and even actin filaments [26], and have been shown to have a large impact on the biological function of proteins and DNA [33, 103]. More recently, several studies have shed light on the relevance of knots also in nanotechnological applications [26, 105, 108, 129, 46, 126].

Physical knots appear and diffuse spontaneously along polymer chains [35, 155, 153, 102, 159, 70], on which they acquire a metastable tightness [66, 48, 49], and can affect structural and dynamical properties like their radius of gyration [53, 113, 119, 95], tensile strength [137, 26], diffusion constants [142, 162], and translocation dynamics through a pore [129].

Most studies up to now focused on the properties of single knots, investigating how knot occurrence probability [81, 120], size [112, 97, 95], and dynamical properties [153, 54] depend on physical characteristics of the system under study, such as polymer thickness [135, 141], confinement [27, 99, 108, 109, 124], stretching force [59, 41, 70, 102], crowding and solution density [76, 131, 57]. Polymers, though, can host multiple knots. These, referred to as *composite knots*, are actually the most probable type of knots in the case of long polymers [144, 55]. The study of composite knots is of great interest since the presence of interactions among their prime components may alter the overall properties of the polymer. For example, knot colocalization on a stretched polymer may significantly diminish its tensile strength with respect to a chain with a series of localized, non-intertwined prime knots.

Since prime knots tied on polymers in solutions or under weak mechanical stretching are weakly localized, their length growing sublinearly with the length of the polymer [59, 112, 97, 95, 41], when tied on the same polymer they are expected, in the thermodynamic limit of infinitely long chains, to be statistically independent. In fact, computational studies have shown that in the thermodynamic limit the characteristic exponent, relating the configurational entropy of a polymer ring to

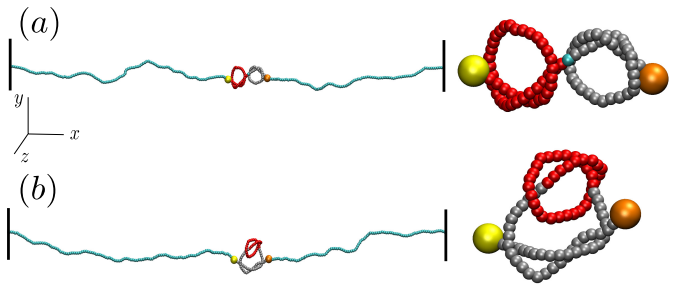


Fig. 4.1 Snapshots from simulations of a chain with bending rigidity $\kappa = 20k_BT$ containing two 3_1 knots of different chirality. The shortest knotted portions of isolated prime components (zoomed in the right part of the figure) are highlighted in red and grey. Yellow and orange beads, whose sizes were artificially increased for clarity, mark the beginning and end of the composite knots, respectively.

its contour length, can be written as in the case of an unknotted ring augmented by the number of knots that are present, and that the knotting probability of composite knots factorizes into that of their prime knot components [30, 150]. However, in finite-size polymers the situation can be quite different. In this case the size of knots is non-negligible and therefore they can not be mapped onto independent point-like decorations [133]. On the contrary, knots can be expected to be intertwined, so that one is inside another, and to interact with each other. Dommersnes *et al.* have shown that knots tied on a short unscreened charged ring become as tight as possible and maximize their relative distances along the ring [56], while Virnau and coworkers, simulating a double-knotted stretched dsDNA chain, have shown that knots can become intertwined in a minimum free energy configuration [149]. Furthermore, a recent study by one of us [151] reported that knot size remains an important factor in freely jointed rings of up to thousands of bonds, where the probability of finding intertwined knots remains significant. These results suggest that finite size effects are relevant in most biological phenomena and nanotechnological applications involving stiff polymers such as dsDNA, and that indeed the overall effect of knots may depend on some complex interaction between them.

To shed further light on the interactions between knots along a finite-size polymer, we consider an idealization of a typical optical tweezers experiment [149, 59, 41] in which a semiflexible double-knotted chain is stretched between two impenetrable walls, describing the effect of tweezer confinement on the terminal polystyrene beads (see Fig. 1). The separation of the chain termini is such that the

knotted polymer is in the high-force stretching regime in which knots are strongly localized, their size showing only small fluctuations around an average size which scales solely with the applied force [59, 41]. Using exhaustive Molecular Dynamics (MD) simulations of polymers containing different pairs of prime knots, we show that not only does a double-knotted dsDNA filament under tension possess a free energy minimum when the two knots are intertwined [149], but also that the depth of this minimum depends on the relative chirality of the two knots. We show that this dependence of the effective interaction on the chirality originates from an interplay between chain entropy and bending energy, which is dominated by the latter.

The dsDNA is modeled as a chain of $N = 300$ beads of diameter σ connected by FENE (finitely extensible nonlinear elastic) bonds [65, 82]. A Weeks-Chandler-Anderson (WCA) potential enforces the excluded volume, and a harmonic bending energy term induces a finite persistence length. The chain is stretched along the X axis, and its termini are kept fixed in correspondence of two repulsive walls, modeled as WCA potentials, which prevent the knots from untying. The total potential energy of the system is thus:

$$U_{tot} = U_{\text{WCA}} + U_{\text{FENE}} + U_{\text{bending}} + U_{\text{walls}} \tag{4.1}$$

where the WCA potential is taken as:

$$\begin{aligned} U_{\text{WCA}} &= \frac{1}{2} \sum_{(i,j),j\neq i}^{N} V(d_{i,j}) \\ V(r) &= \begin{cases} 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + \frac{1}{4} \right] & \text{for } r \leq 2^{1/6}\sigma \\ 0 \ \text{ otherwise} \end{cases} . \end{aligned} \tag{4.2}$$

The WCA strength $\varepsilon = 1 k_B T$ and the characteristic length scale σ are taken as the energy and length units, respectively. All other dimensional quantities are expressed in terms of reduced units defined through ε, σ and the bead unit mass m. Time is measured in the MD time units $\tau_{MD} = \sigma\sqrt{m/\varepsilon} = 1$. The FENE potential reads:

$$U_{\text{FENE}} = -\sum_{i=0}^{N-2} \frac{\kappa_{fene}}{2} \left(\frac{R_0}{\sigma}\right)^2 \ln\left[1 - \left(\frac{|\vec{u}_i|}{R_0}\right)^2\right] \tag{4.3}$$

where $\vec{u}_i \equiv \vec{r}_{i+1} - \vec{r}_i$ is the vector pointing to the bead $i+1$ from bead i, and $|\vec{u}_i|$ is thus the distance of the bead centers i and $i+1$. The values of the maximum bond length $R_0 = 1.5\sigma$ and the FENE interaction strength $\kappa_{fene} = 30\varepsilon$ are the customary ones for the Kremer-Grest model [65]. The harmonic bending potential is taken in its standard form:

$$U_{\text{bending}} = \sum_{i=1}^{N-1} \kappa \left(1 - \frac{\vec{u}_i \cdot \vec{u}_{i+1}}{|\vec{u}_i||\vec{u}_{i+1}|}\right) \tag{4.4}$$

where $\kappa = 20 k_B T$ is the bending stiffness of the chain, inducing a persistence length $l_p = 20\sigma$. Setting $\sigma = 2.5$ nm gives us $l_p = 50$ nm, typical of dsDNA in high monovalent salt concentrations [135].

The potential of Eq. 6.2 is used to perform underdamped MD simulations in an implicit solvent with a Langevin thermostat and time step $\Delta t = 0.01\tau_{MD}$, with the friction self-correlation time $\tau_{frict} = 10^3 \tau_{MD}$.

To assign a topological state to subsections of the chain, we used the *Minimally Interfering Closure* [152]. We define a knotted portion of the chain as the shortest segment featuring a specific knotted topology upon closure according to the Alexander polynomial[1]. With some abuse of language, we will refer to such portions as “knots" in what follows. By applying this procedure to composite knots, we are able to identify both the chain portion hosting the whole composite knot, as well as those hosting its “isolated” prime components. Following ref. [151], we consider a prime component to be isolated when it can be excised, and its ends joined, without at the same time untying the second knot, see Fig 4.1.

Six topologically different setups have been investigated, namely: $(3_1^+ 3_1^+)$, $(3_1^+ 3_1^-)$, $(4_1 3_1^+)$, $(4_1 3_1^-)$, $(5_1^- 3_1^-)$, $(5_1^- 3_1^+)$. The chirality of each prime knot has been established using the writhe, that is, the sum of the signed crossings of the knot in its minimal diagram. $+$ and $-$ superscripts refer to *right* and *left* handedness according to the right-hand rule convention [105, 74]. In our setup, $(++)$ and $(--)$ composite knots are related by a mirror transformation; therefore, we do not attach any importance to the overall chirality of the composite knot, but only on the relative chirality of its prime components. Since no cross-passage is allowed in our simulations, the chirality of all knots is preserved during the simulations. In all setups, the termini of the chain were kept fixed at a distance of $L = 205\sigma$,

[1] We remark that the inability of the Alexander polynomial to distinguish different chiralities does not affect our results, since in our setups the chirality of the knots cannot change and is fixed by the starting configuration.

corresponding to forces of about 1-4 pN at $T = 300K$, depending on the knot complexity, applied on both termini. At these forces, corresponding to a strong stretching regime [41], the knot lengths show relatively small fluctuations around their average values, as reported in Table 4.1. For each of the six topologies under examination, 40 independent simulations were performed, each consisting of an initial equilibration phase of 2×10^7 τ_{MD} and a production phase of 2×10^9 τ_{MD}.

	Separated		Intertwined	
	$l_k^{3_1}$	l_k^{other}	l_k^{incl}	l_k^{comp}
5_1^- # 3_1^-	29.5 ± 3.1	50.5 ± 4.3	30.3 ± 4.7	72.4 ± 3.7
5_1^- # 3_1^+	29.6 ± 3.1	50.6 ± 4.3	33.9 ± 5.2	72.9 ± 3.7
4_1 # 3_1^+	32.9 ± 3.6	44.5 ± 4.0	30.4 ± 3.3	72.6 ± 3.8
4_1 # 3_1^-	32.9 ± 3.7	44.5 ± 4.1	30.4 ± 3.3	72.6 ± 3.8
3_1^+ # 3_1^+	35.5 ± 2.2		33.0 ± 3.7	67.2 ± 4.1
3_1^+ # 3_1^-	35.7 ± 2.3		34.7 ± 3.9	67.0 ± 4.0

Table 4.1 Average knot lengths for different topologies. In the “Separated" columns are reported the knot lengths for the prime components when they are not intertwined. The “Intertwined" columns report the average lengths of the isolated prime component that has been included (labeled l_k^{incl}) and of the whole composite knot in an intertwined configuration (labeled l_k^{comp}). In those cases where the 4_1 is entwined by the 3_1, constituting the 8% of the intertwined configurations for this topology, the size of the 4_1 is 42.5 ± 3.7, and the length of the whole composite knot is 70.1 ± 3.6. We remark that the knots under investigation here are quite tight, with lengths almost half (55%) those of the knots studied in ref. [149] in the case of the $3_1\#4_1$ systems.

As a first case we investigate the $3_1^{\pm}5_1^-$ system, taking into account two different chiralities of the trefoil knot: 3_1^+ and 3_1^-. The fractions of intertwined states in the 3_1^+ and 3_1^- cases are 0.585 and 0.447, respectively. In both setups the largest knot, the 5_1, swells up to let the 3_1 knot in. The frequency with which the trefoil enters or exits the 5_1 is 7.6×10^{-9} τ_{MD}^{-1} for the $5_1^-3_1^-$ $(--)$ pair, and 13.2×10^{-9} τ_{MD}^{-1} for the $5_1^-3_1^+$ $(-+)$ pair [2].

The data in Table 4.1 show that the length of separate prime components is independent of their relative chirality for all topologies under study. This allows us to introduce a collective descriptor, or an *order parameter*, D, defined as the oriented distance between the knot centers. This is measured as the number of

[2]The number of observed events in which the trefoil knot enters into or exits an intertwined state with the 5_1 knot is 580 out of $\sim 7.6 \times 10^{10}$ time steps for the $5_1^-3_1^-$ $(--)$ pair, and 992 out of $\sim 7.5 \times 10^{10}$ time steps for the $5_1^-3_1^+$ $(+-)$ pair.

chain beads from the center of the 5_1 knot to the center of the 3_1 knot, $D = c_{3_1} - c_{5_1}$. A similar definition can be provided for the configurations in which the two knots are intertwined and the knot identification algorithm allows us to identify only one prime component, the one which has been entwined by the other knot. In those cases we identify the center of the swollen knot with the center of the whole composite knot in the expression for D [149]. Therefore, $D = 0$ in those configurations in which the two knots are intertwined and the innermost knot is located exactly at the center of the outermost knot. A schematic representation of this collective descriptor is provided in Fig. 4.2.

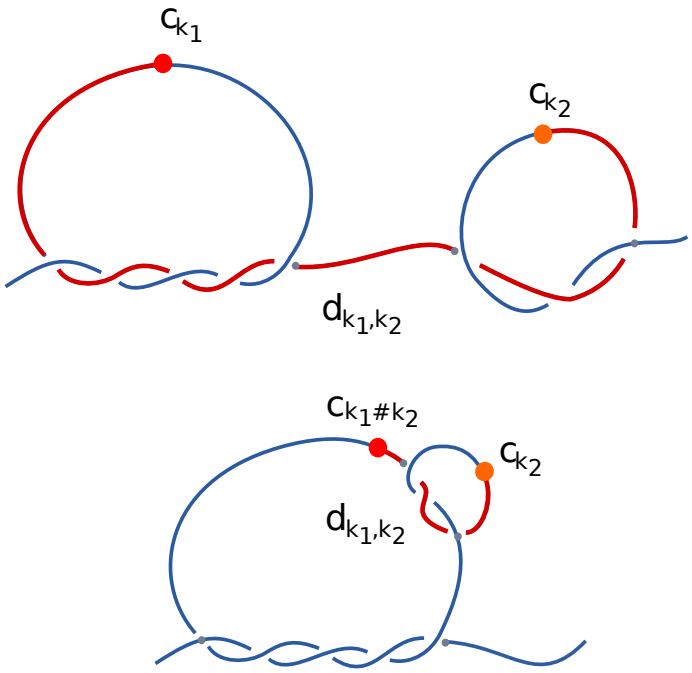


Fig. 4.2 Schematics of the collective order parameter D measuring the linear distance between two prime knots, in this case a 5_1 and a 3_1. When both prime components are isolated, the order parameter is given by $D = c_{k_2} - c_{k_1}$, where $c_{k_i} = (e_{k_i} + s_{k_i})/2$ is the center of knot i on the chain. Here e_{k_i} and s_{k_i} stand for the last and the first bead of the i-th isolated knotted portion. k_1 is always taken to be the most complex knot, in this case the 5_1. When the knots are intertwined, we identify the center of the swollen knot with the center of the whole composite knot, $c_{k_1\#k_2}$.

By counting the relative number of MD configurations for which the knot components are separated by a given distance D we can obtain the probability distribution $P(D)$ and, correspondingly, the free energy $F(D) = -k_B T \ln P(D)$. The latter is reported for the two $3_1^{\pm}5_1^{-}$ systems in Fig. 4.3a.

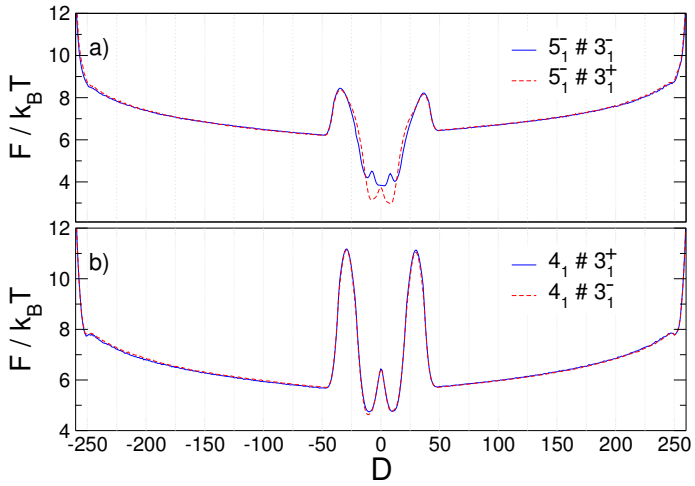


Fig. 4.3 (a) Free energy, $F(D)$, as a function of the linear distance between the knots for $5_1^-3_1^-$ (blue solid line) and $5_1^-3_1^+$ (red dashed line). (b) Same as in (a) for the knots $4_13_1^+$ (blue solid line) and $4_13_1^-$ (red dashed line). Note that in this latter case the two quantities coincide. The free energies for the $4_13_1^\pm$ reported in panel (b) can be decomposed to distinguish the cases in which the 4_1 intakes the 3_1 in and those in which the opposite happens; no substantial difference depending on the relative chirality is to be observed (data not shown).

From the profiles in Fig. 4.3a we observe that $F(D)$ increases with increasing $|D|$, a behaviour that can be intuitively attributed to the entropic cost of placing two knots at large distance on a long, yet finite chain [133]. Consistent also with previous observations [149], for small values of $|D|$ we detect two barriers and two minima in the free energy, corresponding to configurations in which the two knots are intertwined. Most interestingly and unexpectedly, we observe that the depth of these minima depends on the *relative chirality* of the knots. When the two components have opposite chirality, the corresponding free energy minimum is $\sim 1k_BT$ deeper than for the system in which the two chiralities are identical.

This conclusion is reinforced by simulations performed on a $3_1^\pm4_1$ composite knot, where one of the knots, the 4_1, is achiral. The free energy in this case, reported in Fig. 4.3b, does not depend on the chirality of the 3_1 component, as it is indeed expected since there are no other chiral entities in the setup. In the simulations the intertwined states with the 4_1 including the trefoil are the most probable, making for the 92% of observed configurations [3].

[3] For this topology we measure 1280 events in which the one knots enters or exits the other, over a total of $\sim 7.5 \times 10^{10}$ time steps.

To understand if the sole relative chirality of two knots can mark a difference in their preference to stay intertwined or to separate along a chain under tension, we consider a system composed by two otherwise identical trefoil knots. In this case we use as order parameter the absolute value of D, since when the two trefoil knots have the same chirality they become effectively indistinguishable. The free energy profiles, Fig. 4.4a, corroborate that also in this case the system with two knots having opposite chiralities has a lower free energy minimum when the two components are intertwined. We note here that the presence of the two repulsive walls may impact the free energy profiles, since our parameter D does not distinguish whether the knots are near the walls or far from them. In order to rule out possible distortions due to the interactions between knots and walls, the free energy profiles have also been computed excluding all those configurations in which the knots were separated from each wall by a distance lower than $2l_p$. The resulting profiles (data not shown) are perfectly consistent with the ones reported in Figs. 4.3 and 4.4, thus validating the robustness of the observed behavior.

Further insight into the cause of this effect can be obtained by analyzing the size of the intertwined configurations. As reported in Table 4.1, we find that for all investigated topologies the length of the composite knot in the intertwined state, l_{comp}, does not depend on the relative chirality of the prime components. Interestingly, we observe that both for the $3_1^{\pm}5_1^{-}$ and for the $3_1^{\pm}3_1^{-}$ systems the length of the nested, isolated prime components in the intertwined state is slightly larger in the $(+-)$ case [4].

We proceed to separate the free energy $F(D)$ in its energetic and entropic components, by first computing the average internal (potential) energy of the configurations, and subsequently obtaining the entropy through the standard relation $F = E_{int} - TS$. The results, reported in Fig. 4.4a-c, show two interesting features. First, they confirm that the observed increase of F with $|D|$ when the two knots are separated is purely entropic. Secondly, and more importantly, they show that the differences we observe in the free energies of the $(+-)$ and $(++)$ systems originate from a complex interplay of internal energy and entropy. Specifically, the entropic contribution is higher for the $(++)$ case but is not high enough to overcome the energetic contribution favoring the $(+-)$ knot. The potential energy can be further

[4] In the case of the $(+-)$ topology, we observe 3720 events out of $\sim 8.0 \times 10^{10}$ time steps in which one trefoil enters or exits the other; for the $(--)$ topology we have 3108 such events out of $\sim 8.0 \times 10^{10}$ time steps.

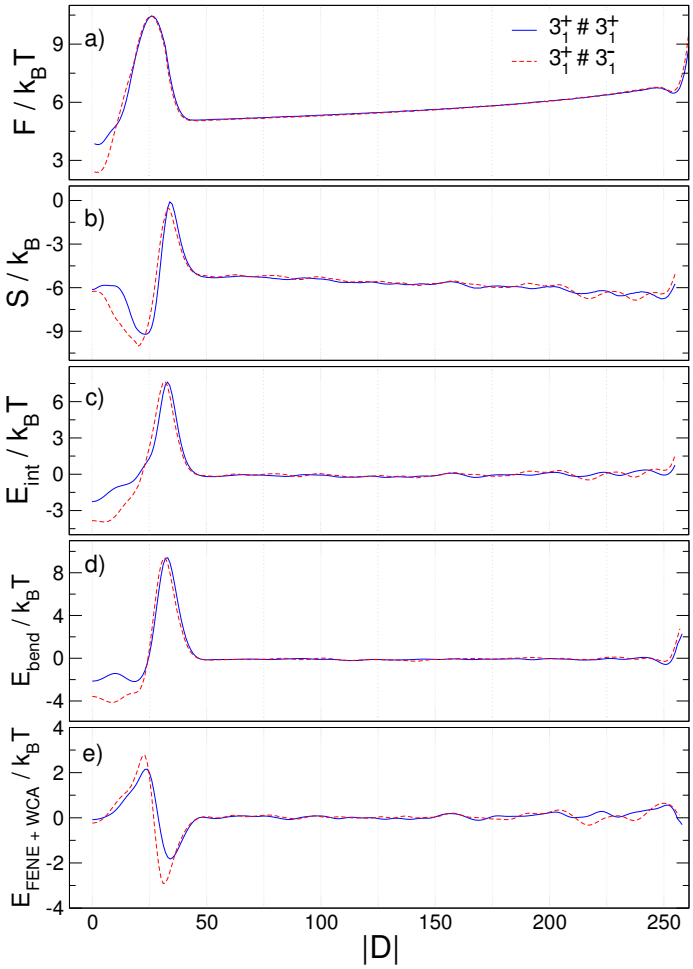


Fig. 4.4 Free energy of the $3_1 3_1$ system with same and opposite relative chirality, as a function of the separation D between the two knots, decomposed in their different energetic and entropic contributions. We report the free energy $F(D)$ (a), the entropy S (b), and the internal potential energy E_{int} (c). The entropy is obtained through the relation $F = E_{int} - TS$. In panels (d) and (e) we show how the internal potential energy is distributed, respectively, between the bending term and the remaining two contributions, namely the FENE and WCA potentials. See the main text for a discussion of the results.

decomposed into its main components: steric hindrance, bond extension, and bending energy. The data presented in Fig. 4.4d-e clearly show that while all other energetic contributions are similar, the bending energies of the $(+-)$ and $(++)$ systems differ significantly at the position of the minimum of $F(D)$ by the same amount, $\sim 1.5 k_B T$. The same qualitative result holds also for the $3_1^{\pm} 5_1^{-}$ topologies (data not shown).

Given the observed competition between bending energy and entropy in the intertwined state, with the $(+-)$ system showing a lower bending energy but losing more entropy than the $(++)$ system, it is tempting to ascribe the difference in their free energy profile to a significantly different arrangement of the nested knot within the hosting knot in the two setups. However, further analyses and simulations are required to elucidate the exact mechanics underlying the chiral contribution, which is the object of an ongoing study.

Summing up, we have shown that a double knotted semiflexible polymer chain under strong stretching possesses a free energy minimum when the two knots are intertwined, showing that relatively tight knots can still pass through each other, and also that the depth of this minimum depends on the relative chirality of the two knots. In order to observe this effect, both knots must be chiral, with the knots of opposite chirality displaying a higher preference to stay intertwined. Furthermore, we showed that the major player in the chirality effect is the bending energy of the chain, which, we recall, is set to the characteristic persistence length of dsDNA.

The question then emerges, as to what are the specific features of the bending energy that would engender a chirality effect in knot interaction, and whether the stretching of the chain enters in some way as a significant constraint. If the chirality effect, described above, turns out to be robust and observable in a broad range of parameters, one can speculate further as to its importance in particular for chiral polymers such as dsDNA.

Acknowledgments

L. Tubiana and R. Podgornik acknowledge support from the Slovenian Agency for Research and Development (ARRS grant No J1-4134). L. Tubiana acknowledges also support from the Mahlke-Oberman Stiftung and the European Union's Seventh Framework Programme for research, technological development and demonstration (grant No 609431). S. Najafi and R. Potestio are thankful to M. Heidari and R. Menichetti for an attentive reading of the manuscript and useful comments.

Chapter 5

Role of bending energy and knot chirality in knot distribution and their effective interaction along stretched semiflexible polymers

This Chapter has been published as a research paper in Polymers journal. It is reprinted here with permission from the publisher.

Saeed Najafi, Rudolf Podgornik, Raffaello Potestio, and Luca Tubiana
Role of Bending Energy and Knot Chirality in Knot Distribution and Their Effective Interaction along Stretched Semiflexible Polymers
Polymers 8, 347 (2016)

Abstract

Knots appear frequently in semiflexible (bio)polymers, including double-stranded DNA, and their presence can affect the polymer's physical and functional properties. In particular, it is possible and indeed often the case that multiple knots appear on a single chain, with effects which have only come under scrutiny in the last few years. In this manuscript, we study the interaction of two knots on a stretched

semiflexible polymer, expanding some recent results on the topic. Specifically, we consider an idealization of a typical optical tweezers experiment and show how the bending rigidity of the chainand consequently its persistence lengthinfluences the distribution of the entanglements; possibly more importantly, we observe and report how the relative chirality of the otherwise identical knots substantially modifies their interaction. We analyze the free energy of the chain and extract the effective interactions between embedded knots, rationalizing some of their pertinent features by means of simple effective models. We believe the salient aspect of the knot–knot interactions emerging from our study will be present in a large number of semiflexible polymers under tension, with important consequences for the characterization and manipulation of these systems—be they artificial or biological in origin—and for their technological application.

5.1 Introduction

A general comprehension of the statistical behavior of semiflexible polymers strongly stretched by an externally imposed tension is relevant to understanding the details of DNA processing in cells [121], as well as to quantify their behavior in single molecule experiments [100]. In the latter, DNA is standardly prepared in a linear, topologically trivial form; on the contrary, the vagaries of the intracellular [37, 103, 118, 33] and viral DNA environment [28, 27, 99, 98] suggest that we also need to focus our attention on more complicated, knotted forms of DNA [33, 52, 63]. In this context, the behavior of single isolated knots on DNA has received plenty of attention [26, 32]; in particular regarding the size, distribution, and dynamics of different prime knots [35, 59, 155, 159, 70, 102, 153, 124, 66, 48, 49, 41], while more probable [144] and experimentally observed [58] composite knots—topological entanglements made of multiple prime knots—have remained much less scrutinized.

When multiple knots are present on a finite-size polymer, they do not behave like independent point-like decorations, as theorized for infinitely long polymer rings [30, 150]. Instead, they show a propensity to either intertwine or repel each other, which is controlled by the polymer size [151], its bending energy [149] and its electrostatic self-repulsion [56]. These features suggest that finite size effects are relevant in most biological phenomena and nanotechnological applications

involving stiff polymers such as double-stranded DNA (dsDNA), and that indeed the overall effect of the presence of knots on the chain may depend on some complex interaction between them.

The questions we will address in what follows are connected with the details of the free energy pertaining to composite knots tied on a semiflexible polymer, which was shown to exhibit a minimum for intertwined knots [149]. Specifically, we will be concerned with the nature of the effective interaction between prime components of the composite knots that was recently discovered to exhibit unexpected features: apart from depending on the bending stiffness of the underlying polymer, it was shown to depend also on the *relative chirality* of the interacting knots [115]. In fact, in our recent work [115], we found that not only do two knots on a stretched semiflexible polymer under external tension present a free energy minimum when intertwined, but, more significantly, that the depth of this minimum depends on the relative chirality of the two knots. This minimum is more pronounced for knots of opposite chiralities, and the chirality-dependent features of the interaction are present even for otherwise identical trefoil knots.

The chirality effect originates from a complicated dependence of the bending energy on the configuration of the interacting knots, which, within the intertwined knots, is arguably caused by a different arrangement of the inner (nested) knot. As a corollary, this finding implies that torus knots with opposite chiralities should remain intertwined longer than knots with the same chirality—a fundamental feature of the interaction between knots that should change our perspective on their overall importance and salient features.

In order to characterize the role of chirality and chain stiffness on the properties of the composite knots, we simulate a typical optical tweezers experiment involving a knotted polymer [59, 149, 41]. In particular, we consider two different topologies: A $3_1^+3_1^+$ and a $3_1^+3_1^-$ knot, distinguished only by the relative chiralities of their components, for different degrees of polymer stiffness. By means of molecular dynamics (MD) simulations in implicit solvent with an underdamped Langevin thermostat, we show explicitly that indeed the effects of knot chirality become more evident for larger chain stiffness.

The behavior of the two knots is described in terms of their linear distance along the chain, and through the analysis of the system's free energy (expressed in terms of this measure), we are able to identify three different regimes. In the first regime,

the knots are separated along the polymer and the free energy is dominated by an entropic interaction between them. The second regime appears when the knots are close to each other but still not intertwined, and their interaction is dominated by the steric repulsion of the loops. The third and most intriguing regime emerges when the two knots are intertwined: in it, the free energy of the system depends on both the bending energy of the polymer and on the relative chirality of the knots.

These three regimes are rationalized by means of simple, mechanistic models based on the various contributions of different knotted configurations to the total entropy and/or elastic energy of the chain. In spite of their crudeness, these models illuminate the nature of chirality-dependent interactions between knots in stretched stiff polymers, and provide a solid starting point for further theoretical developments, as well as for the design of systems in vivo or in silico with specific, tailored physical properties.

5.2 Materials and Methods

5.2.1 Model and Simulation Methodology

We model a single stiff linear polymer chain under external tension as a sequence of $N = 300$ spherical beads attached to and stretched between two impenetrable planar hard walls, approximating the polystyrene beads used in a typical optical tweezer experiment [39]. The polymer excluded volume interaction is accounted for by a purely repulsive WCA potential between each pair of beads, while the connectivity of the chain is described by FENE bonds [65, 82]. The bending stiffness of the polymer chain enters through an elastic filament curvature deformation energy depending on the relative angle between two successive links along the chain.

The total potential energy of the system can therefore be expressed as a sum of four components:

$$U_{\text{tot}} = U_{\text{WCA}} + U_{\text{FENE}} + U_{\text{bending}} + U_{\text{walls}} \tag{5.1}$$

where the WCA potential is taken as:

$$U_{\mathrm{WCA}} = \frac{1}{2} \sum_{(i,j),j\neq i}^{N} V(r_{i,j}) \tag{5.2}$$

with

$$V(r) = \begin{cases} 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + \frac{1}{4} \right] & \text{for } r \leq 2^{1/6}\sigma \\ 0 \quad \text{otherwise} \end{cases} . \tag{5.3}$$

The WCA interaction strength $\varepsilon = 1 k_B T$ and the characteristic length scale σ are taken as the energy and length units, respectively. All other dimensional quantities are expressed in terms of reduced units defined through ε, σ, and the bead unit mass m. Time is measured in the MD time units $\tau_{\mathrm{MD}} = \sigma\sqrt{m/\varepsilon} = 1$. The FENE potential is given by:

$$U_{\mathrm{FENE}} = -\sum_{i=1}^{N-1} \frac{\kappa_{\mathrm{FENE}}}{2} \left(\frac{R_0}{\sigma}\right)^2 \ln\left[1 - \left(\frac{|\mathbf{u}_i|}{R_0}\right)^2\right] \tag{5.4}$$

where $\mathbf{u}_i \equiv \mathbf{r}_{i+1} - \mathbf{r}_i$ is the connecting vector between bead $i+1$ and bead i (directed towards the former), and the modulus $|\mathbf{u}_i|$ is thus the separation between the centers of beads i and $i+1$. The value of the maximum bond length $R_0 = 1.5\sigma$ and the FENE interaction strength $\kappa_{\mathrm{FENE}} = 30\varepsilon$ are the customary values pertinent to the Kremer–Grest model [65]. Finally, the elastic bending potential is taken in the form:

$$U_{\mathrm{bending}} = \sum_{i=1}^{N-1} \kappa_b \left(1 - \frac{\mathbf{u}_i \cdot \mathbf{u}_{i+1}}{|\mathbf{u}_i||\mathbf{u}_{i+1}|}\right) \tag{5.5}$$

where κ_b is the bending rigidity of the chain.

The full interaction potential of Equation (6.2) is employed in underdamped MD simulations in an implicit solvent. The simulation time step is $\Delta t = 0.01\tau_{\mathrm{MD}}$, and the friction self-correlation time is $\tau_{\mathrm{frict}} = 10^3\tau_{\mathrm{MD}}$. For each system under examination, we run 30 independent simulations, each covering $2 \times 10^7 \tau_{\mathrm{MD}}$, with an initial equilibration phase of length $2 \times 10^5 \tau_{\mathrm{MD}}$. The simulations were performed making use of the LAMMPS [122] MD package.

The confining hard walls and the effective impenetrability of the chain bonds—provided by the combination of the FENE potential and the WCA interaction—ensure that the topology of the system remains fixed to the one set by the initial

configuration of two trefoil knots; thus, either $3_1^+3_1^-$ $(+-)$ or $3_1^+3_1^+$ $(++)$. The $+$ and $-$ superscripts indicate the handedness of the knots, right or left, according to the sum of the signed crossings in their minimal diagrams [105, 74]. Since in our setup $(++)$ and $(--)$ knots are related by a mirror transformation, their physical properties are equivalent, and we only simulate the former.

To study the dependence of the chirality effects on the bending energy, we simulate each of the two setups with different chirality at various values of the chain stiffness; namely, $\kappa_b = 2.5$ to 20.0 k_BT. In all setups, the termini of the chain are kept fixed at a distance of $L = 205\sigma$, corresponding to stretching forces of about 1–7 pN at $T = 300$ K, depending on the bending rigidity of the chain, and applied at both termini.

In order to study the statistical properties of the knots on the polymer chain, we need to identify which portions of the chain host the knots. The identification of the smallest knotted portion of the chain (that is, the segment that we define as the knot) is enabled by the usage of the Minimally Interfering Closure [152] to circularize open subsections of the chain into auxiliary arcs, whose topology is then established by means of the Alexander determinants in -1 and -2. It is worth note that the inability of the Alexander polynomial to distinguish different chiralities does not affect our results, since we are interested only in the distance between the two trefoil knots. We define a knotted portion of the chain as the shortest segment featuring a specific knotted topology. With some abuse of language, we will refer to such portions as "knots" in what follows. By applying this procedure to composite knots, we are able to identify both the chain portions hosting the whole composite knot, as well as those hosting its "isolated" prime components. Following Ref. [151], we consider a prime component to be isolated when it can be excised, and its ends joined, without at the same time untying the second knot (as depicted in Figure 5.1).

5.3 Results

We investigate the behavior of two trefoil knots on semiflexible chains of different bending rigidity, stretched between two impenetrable walls kept at fixed distance. The investigated bending rigidities span the range from $\kappa_b = 2.5$ to $\kappa_b = 20\ k_BT$, corresponding to different stretching forces, all in the strong stretching regime

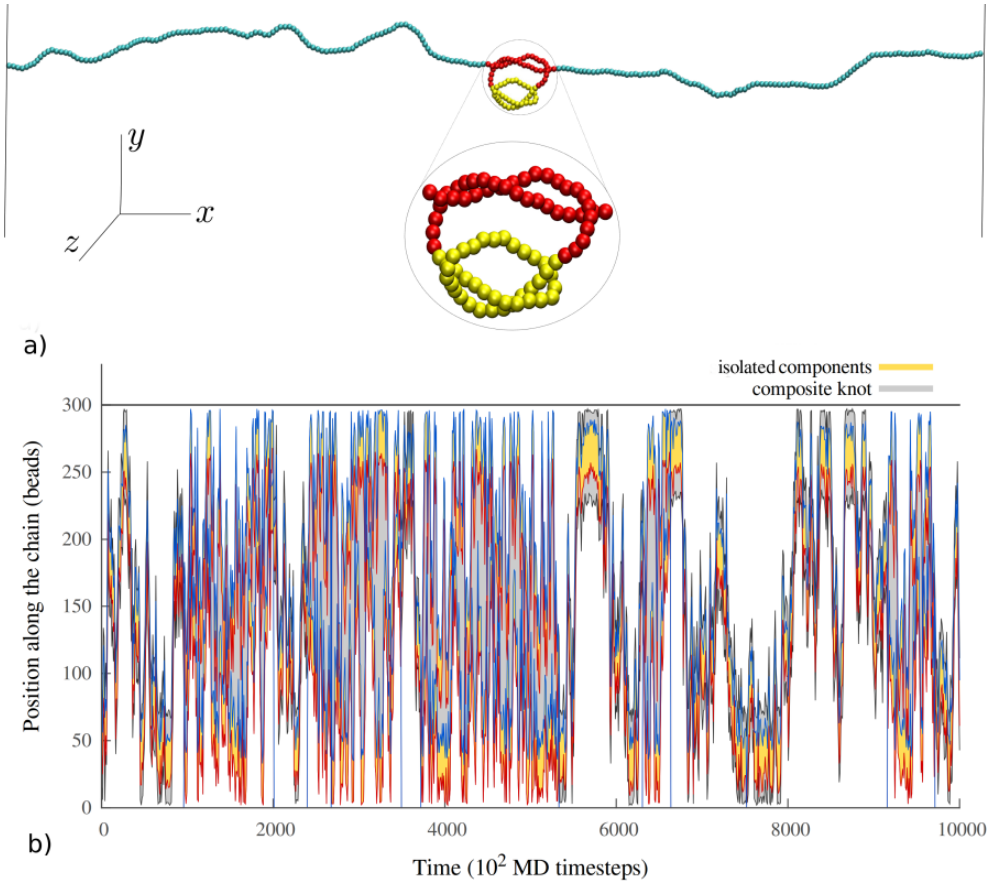


Fig. 5.1 (**a**) An example of a configuration for the (++) system, with both knots intertwined. In this configuration, only one prime component (marked in yellow) is identified as "isolated" by our knot identification scheme based on the Minimally Interfering Closure [152]; (**b**) A small portion of an MD trajectory for the (++) system. Yellow shaded regions indicate the portions of the chain occupied by the isolated prime knots, and red and blue lines indicate their first and last bead, respectively. The portion of the chain taken up by the whole composite knot is reported as a gray shaded area. When the knots are intertwined, our algorithm identifies only one isolated knot, which can be seen in the trajectory as a single yellow region surrounded by gray boundaries. Note that the intertwined composite knot can still travel from one side of the chain to the other.

(see Methods). For each value of κ_b, we ran 30 independent simulations of about $10^7 \tau_{\mathrm{MD}}$ steps each. We sampled the system by storing the whole polymer configuration every $100\tau_{\mathrm{MD}}$. A typical portion of an MD trajectory for $\kappa_b = 20\, k_B T$ is reported in Figure 5.1b.

The knot localization scheme described in the Methods section allows us to easily distinguish configurations in which the knots are intertwined; i.e., when one

knot is inside the other, from configurations in which they are separated along the chain. These states can be distinguished from the number of isolated components identified by our knot localization algorithm: two, when the knots are separated, and one—the nested knot—when the knots are intertwined. At $\kappa_b = 2.5\ k_BT$, we do not observe any crossing event in which the two knots would become intertwined. Therefore, this value of κ_b is considered in the following only in relation to the properties of the knots when they are separated along the polymer chain. At $\kappa_b = 5\ k_BT$ we observe $\approx$30 crossing events in which the two knots get either intertwined or separated. The number greatly increases with κ_b, to reach $\approx$2500 events at $\kappa_b = 20\ k_BT$, for a simulation of about $\sim 5.6 \times 10^8\, \tau_{\mathrm{MD}}$.

5.3.1 Knot Sizes

The first observable on which we direct our attention is the size of the knots, measured as the number of beads included in the entanglements—both in their separated and intertwined states—in dependence of the bending stiffness κ_b. From Figure 5.2a, we observe that the size of separated knots grows sublinearly with increasing bending rigidity. This is consistent with a simple minimization of the bending energy stored in the loop and braids of the knot, compounded with the effective chain shortening as some of its length is used up by the knots [29, 73]. More interestingly, it appears that the size of the composite, intertwined knot, (shown in Figure 5.2b) depends on the bending rigidity in a non-linear, possibly also non-monotonic fashion, with the knot size deviating substantially from a linear dependence for small values of κ_b. This may be related with recent observations by Poier [124] and Caraglio [41], and points to the importance of the entropic effects that become more important as the bending rigidity is diminished.

Neither the size of the separated knots nor that of the whole intertwined knots manifest a dependence on the relative chirality. This is not the case for the size of the nested knot, as can be appreciated from Figure 5.2b. Here, in contrast with the whole intertwined knot, we have a substantially linear growth of the nested knot size with κ_b, and the sizes of the nested knots for the different chirality—initially having the same value, $\kappa_b = 5\ k_BT$—increase with different rates. Specifically, the nested knot in the setup where both prime components have the same chirality shows a smaller size with respect to the $(+-)$ case. Albeit small, this discrepancy

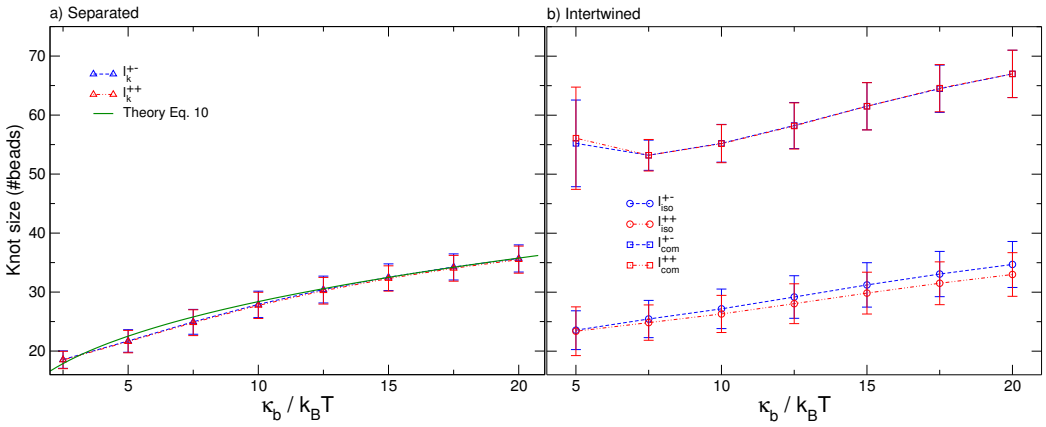


Fig. 5.2 (**a**) Size of separated trefoil knots, plotted as a function of κ_b; (**b**) Size of a nested trefoil knot (circles) as well as of the composite knot (squares) when the two trefoils are intertwined, plotted as a function of κ_b.

points to distinguishably different organizations of the intertwined knots depending on their relative chirality, and we substantiate this expectation in the following.

5.3.2 Knot Free Energy

We proceed to study how the free energy of the $(+-)$ and $(++)$ configurations depends on the different arrangements of the knots along the chain. To do so, we investigate the free energy as a function of a collective order parameter $|D|$, defined as the absolute linear distance between the centers of the two knots [149], see Figure 5.3. Specifically, we identify the starting and ending beads s_i and e_i of both knots on the chain, where the index i pertains to the knot; the center of knot i along the chain is then at $c_i = (s_i + e_i)/2$. When only one trefoil can be identified by the Minimally Interfering Closure (see Methods section), the knots are intertwined and we can take the starting and ending beads of the non-isolated knot to correspond to those of the whole composite knot. Consequently, the collective order parameter can be introduced as the absolute linear distance between the centers of the two knots, defined as

$$|D| = \begin{cases} |c_1 - c_2| & \text{(separated knots)} \\ |c_1 - c_{1,2}| & \text{(intertwined knots)} \end{cases} \tag{5.6}$$

Starting from the definition of $|D|$, we evaluated the free energy $F(|D|) = -k_B T \log(\omega(|D|; N, \langle l_k \rangle))$. The free energy profiles $F(|D|)$ for the $(++)$ and $(+-)$

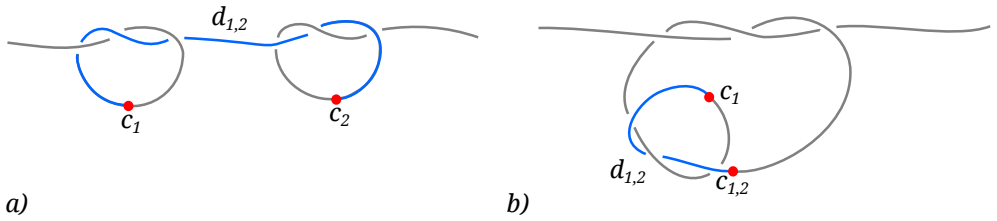


Fig. 5.3 Schematics of the collective order parameter $|D|$ measuring the linear distance between two prime knots. When both prime components are isolated, the order parameter is given by $|D| = |d_{1,2}| = |c_1 - c_2|$, where $c_i = (e_i + s_i)/2$ is the center of knot i on the chain. Here e_i and s_i stand for the last and the first bead of the i-th isolated prime knot. When the two knots are intertwined, the center of the hosting knot is taken to coincide with the center of the composite knot $c_{1,2}$.

chirality configurations at various values of the polymer bending rigidity are reported in Figure 5.4. Consistent with previous results [115], we observe that the intertwined state with $|D| = 0$ becomes a global minimum only when the bending rigidity κ_b reaches large enough values. We observe that such values differ for knots with the same or different chiralities. In addition, due to the high stretching of the polymer chain, the intertwined state of either chirality configuration becomes a global minimum only when $\kappa_b > 10\ k_B T$, a higher value than the one observed in Ref. [149] for a more complex setup ($3_1 4_1$), but at lower stretching forces.

The plots of Figure 5.4 also indicate that the $(+-)$ system always has a deeper minimum than the $(++)$ system, while the rest of the curves remain effectively universal and independent of the bending rigidity, with a slow growth of the free energy with increasing $|D|$. This universal behavior is eventually followed by a small minimum for very high separations of the two knots, corresponding to states in which each knot is localized in the vicinity of a confining wall. Increasing $|D|$ even further, the free energy monotonically increases with a slope that is dependent on the bending rigidity, up to a point at which $|D|$ is close to the maximum possible separation between the knots. The different behaviors of $F(|D|)$ for large values of $|D|$ are due to finite size effects. In particular, the minima and barriers at $|D|$ close to $N = 300$ are caused by the presence of the impenetrable walls, while the universal slow growth of $F(|D|)$ originates from the finite size of the chain itself and can arguably also be observed on unstretched rings [151]. Similar results (excluding the minima and the barriers due to the presence of the walls) were obtained by excluding all configurations in which one knot was closer to the wall than $\langle l_k \rangle$ (data not shown).

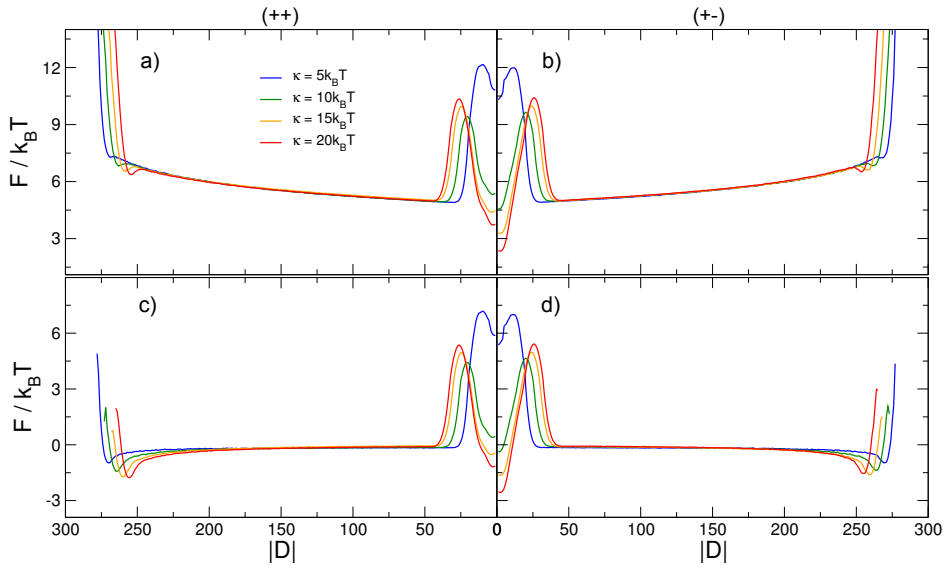


Fig. 5.4 Free energy F as a function of the collective order parameter $|D|$, i.e., the absolute linear distance between the centers of the two knots. $F(|D|)$ for different values of the chain bending rigidity κ_b is reported for (**a**) two trefoil knots with the same relative chirality and (**b**) two trefoil knots with opposite relative chiralities. In (**c**,**d**) we show the same free energies, but with subtracted entropic contribution $-kS(|D|)$, as defined in Equation (5.8).

The universal bending rigidity-independent part of the free energy as a function of the order parameter $|D|$—in the range of values between the intertwined knot state and the knot localization at the boundaries of the chain—can be understood with simple scaling arguments. Assuming that the length of the a trefoil knot (l_k) does not fluctuate significantly from its average value $\langle l_k \rangle$ when the knots are separated, we can map the polymer chain onto a linear string of N beads, along which two chosen segments of length $s = \langle l_k \rangle$ (representing the knots) can slide freely. From the definition of $|D|$, we can see that the two knots of equal length separated by $|D|$ beads take up a portion of the chain of size $\alpha = |D| + \langle l_k \rangle$. An explicit counting of the available microstates shows that there are $N - \langle l_k \rangle + 1 - |D|$ configurations in which the segments are separated by a distance $|D|$ between their centers, without any double counting, since the knots are indistinguishable when separate. The fraction of microstates $\omega(|D|; N, \langle l_k \rangle)$ can then be written as:

$$\begin{aligned} \omega(|D|; N, \langle l_k \rangle) &= \frac{(N - \langle l_k \rangle + 1 - |D|)}{\sum_{|D|=0}^{N-\langle l_k \rangle+1} (N - \langle l_k \rangle + 1 - |D|)} \\ &= \frac{2(N - \langle l_k \rangle + 1 - |D|)}{(N - \langle l_k \rangle + 2)(N - \langle l_k \rangle + 1)} \end{aligned} \tag{5.7}$$

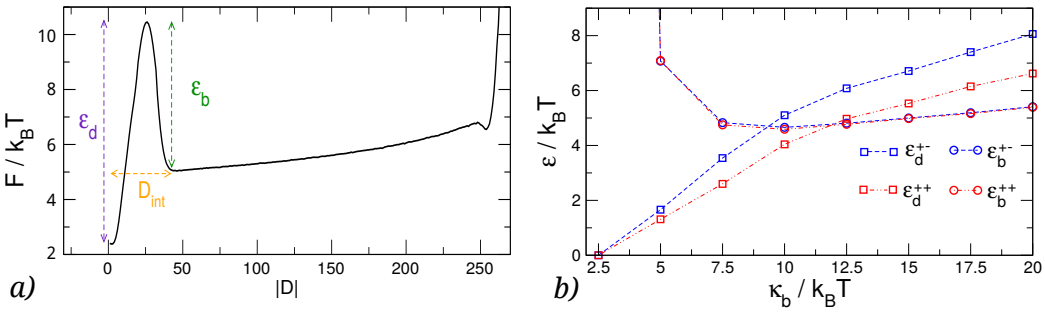


Fig. 5.5 (**a**) Graphical representation of three quantities discussed in the text: the height of the barrier, ε_b, the depth of the minimum, ε_d, and the interaction distance, D_{int}; (**b**) Dependence of the height of the barrier and the depth of the free energy minimum on the polymer bending rigidity, κ_b.

and the entropy as a function of $|D|$ reads:

$$S(|D|;N,\langle l_k\rangle) = \ln \omega(|D|;N,\langle l_k\rangle). \tag{5.8}$$

Subtracting this purely entropic contribution from the free energies in Figure 5.4a,b and making use of the values of $\langle l_k(\kappa_b)\rangle$ reported in Figure 5.2, we obtain a completely flattened-out free energy for $|D|$ between the intertwined knot state and the knot localization at the boundaries of the system. This simple transformation of the free energy accentuates those non-universal features that depend on the bending rigidity, and consequently on the interactions between the two knots or between the knots and the confining walls; see Figure 5.4c,d.

From the plots in Figure 5.4a,b, we extract the ΔF between the minimum and the barrier (ε_d), the height of the barrier with respect to the entropic plateau (ε_b), and the distance D_{int} at which the knots begin to interact. The latter is defined as the value of $|D|$ at which the $F(|D|)$ starts to increase when $|D|$ is reduced below $|D| = N/2$, see Figure 5.5a. The results for ΔF, reported in Figure 5.5b, show that the height of the barrier (ε_b), does not depend on the relative chirality of the two knots, but displays an interesting dependence on κ_b. In fact, ε_b appears to have a very broad minimum for $\kappa_b \sim 10\, k_B T$. On the other hand, we observe that the depth of the minimum (ε_d), increases monotonically with κ_b, and that the separation between the ε_d curves for $(++)$ and $(+-)$ knots increases as well, confirming that the emergence of the chiral effect is finally triggered by the bending rigidity of the chain.

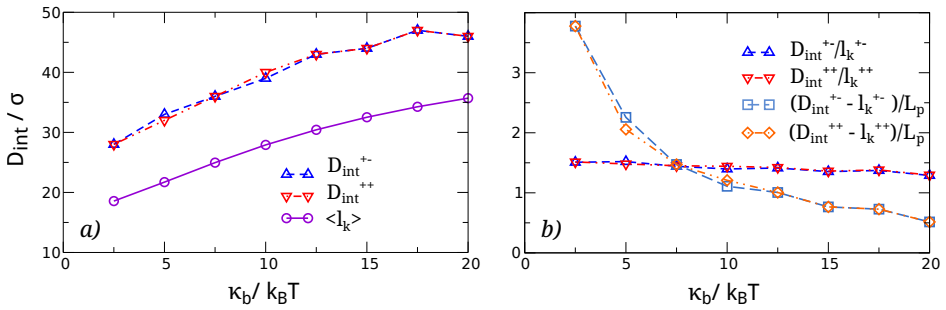


Fig. 5.6 (**a**) Interaction distances of separated knots as a function of κ_b. For comparison, we also report the knot size averaged between the $(++)$ and $(+-)$ systems, $\langle l_k \rangle$. Note that the interaction distance is always larger than $\langle l_k \rangle$; (**b**) Two different rescaling of D_{int}: D_{int}/l_k and $(D_{\text{int}} - l_k)/L_p$, with L_p being the persistence length of the chain.

In light of these results, one is naturally led to wonder if the bending energy of the chain could be implicated in mediating an effective interaction between the knots when they are separated as well. To investigate this aspect, we identified the interaction distance between two well-separated knots. From the data reported in Figure 5.6a, we can see that indeed the knots begin to interact at distances $|D| > \langle l_k \rangle$, the characteristic distance at which their ends along the chain coincide. In order to see if this interaction is dictated by the curvature of the chain segment connecting the two knots and its fluctuations, or by the steric interaction between the two knot loops, we rescale D_{int} by $\langle l_k(\kappa_b) \rangle$. The results reported in Figure 5.6b show that D_{int} increases slightly slower than l_k with k. Since l_k grows sublinearly with the persistence length $L_p \sim \sigma \kappa_b / k_B T$, one can immediately see that $\frac{D_{\text{int}}}{L_p}$ decreases faster than $\frac{D_{\text{int}}}{l_k}$ with κ_b. The same goes for the length of the segment connecting the two knots, $D_{\text{int}} - l_k$, as can be seen from Figure 5.6b. From this analysis, we conclude that the interaction between two simple separate knots on a stretched semiflexible polymer of bending rigidity up to $L_p \sim 20\sigma$ is dictated primarily by the steric hindrance of their loops.

5.3.3 Relative Orientation of the Knots

An important descriptor of the organization of the knots is their relative orientation when they are separated, as well as when they are nested. We thus introduce the knot orientation director $\mathbf{U}_k$, defined as the sum of the vector products of consecutive bond vectors; the sum is extended to all bond vectors contained in the

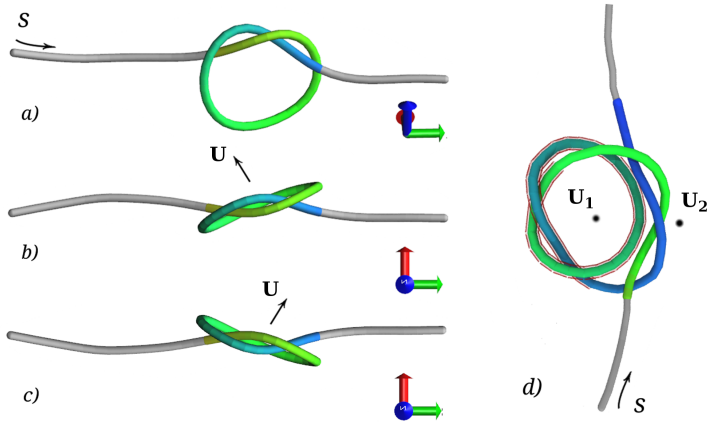


Fig. 5.7 Left-handed 3_1 knot configuration obtained by minimizing the energy viewed from the side (**a**) and from the top (**b**); (**c**) Right-handed 3_1 knot viewed from above. x indicates the pulling direction. In (**a**), the direction of the arc length s is indicated by an arrow, and in all panels by the coloring of the knots, from green to blue for increasing s. In panels (**b**,**c**), we reported the direction of the vector $\mathbf{U}_k$ defined in Equation (5.9). Note that $\mathbf{U}_k \cdot x$ changes sign with the handedness of the knot; (**d**) A $(+-)$ intertwined composite knot, with the nested knot highlighted by a red shading. Note that when the loop of the nested knot lies inside the loop of the outer knot, both orientation directors point out of the page. On the other hand, when the loop of the two knots form an eight, the two directors point in opposite directions.

knot:

$$\mathbf{U}_k = \frac{\sum_{j\in k} \mathbf{u}_j \times \mathbf{u}_{j+1}}{|\sum_j \mathbf{u}_j \times \mathbf{u}_{j+1}|}. \tag{5.9}$$

The direction of $\mathbf{U}_k$ is defined by the arc length along the knot k, according to the right hand rule. In configurations obtained by minimizing the energy of the knot, $\mathbf{U}_k$ identifies the normal to the plane passing trough the knot loop, as depicted in Figure 5.7. Note that knots of different handedness have opposite projections of $\mathbf{U}_k$ along the pulling direction X.

Using the knot orientation director, we introduce two order parameters, which are based on idealizing the knot loops as rigid discs. The first such parameter, θ, captures the (instantaneous) angle between two knot loops, and is defined as:

$$\theta = \arccos(\mathbf{U}_1 \cdot \mathbf{U}_2). \tag{5.10}$$

Since θ measures only the aperture of the cone identified by $\mathbf{U}_1$ and $\mathbf{U}_2$, we consider another parameter to fully capture the relative orientation of the two knots, $\theta_\perp$, which captures the relative rotation of the knot loops with respect to the stretching axis X:

$$\theta_\perp = arctan2\,(sgn\,|\mathbf{V}_1 \times \mathbf{V}_2|, \mathbf{V}_1 \cdot \mathbf{V}_2) \tag{5.11}$$

with:

$$\mathbf{V}_i \equiv \frac{(0, U_i^y, U_i^z)}{\sqrt{(U_i^y)^2 + (U_i^z)^2}}$$
$$sgn = sign([\mathbf{U}_1 \times \mathbf{U}_2]^x)$$

In words, $\theta_\perp$ measures the angle between the projections of the $\mathbf{U}$ vectors on the YZ plane (that is, the plane perpendicular to the stretching direction). These projections constitute the normalized two-dimensional vectors $\mathbf{V}$. The sign of the angle, which can be determined by means of the $arctan2$ function (analogous to the two-argument `FORTRAN` routine `ATAN2` for the inverse tangent), is determined by the X component of the vector product between the full $\mathbf{U}$ vectors: if the resulting vector points in the same direction as X, the sign is positive.

When the two knots are intertwined, $\mathbf{U}_1 = \mathbf{U}_{\text{nested}}$ is defined for the bonds of the nested knot, while $\mathbf{U}_2$ is defined for the bonds pertaining to the region of the composite knot that is complementary to the nested knot. This allows us to use the same orientational parameters both when the knots are separated, and when they are nested.

We first study the relative orientation of the knots when they are separated. From the data in Figure 5.8a, we notice that the projection of the free energy along θ presents broad minima for angles close to $\pi/4$ for the $(++)$ system and angles close to $3\pi/4$ for the $(+-)$ system. The minimum energy angles θ^* for the two systems are symmetric with respect to $\pi/2$; i.e., $\theta^*_{(++)} = \pi - \theta^*_{(+-)}$. On the other hand, it is clear from Figure 5.8b that there is no preferred relative orientation of the two loops on the YZ plane. The minima of $F(\theta)$ thus simply capture the different alignment of the braids of the two knots, which depends on the chirality of the knots, as shown in Figure 5.7.

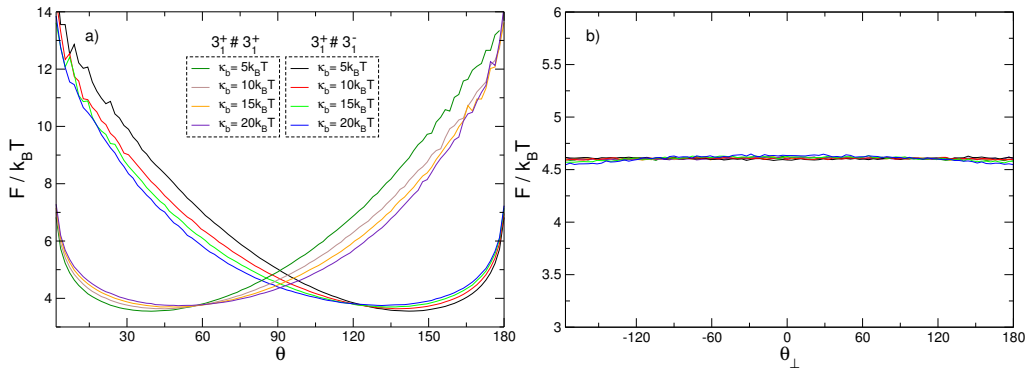


Fig. 5.8 Free energy of two separated trefoil knots projected along (**a**) θ and (**b**) $\theta_\perp$ for four different values of polymer bending rigidity.

When the knots are intertwined, we expect that their arrangement will be sensitive to the their relative chirality. Since the difference between the depth of the minima of the $(++)$ and $(+-)$ systems increases with increasing bending rigidity, we further expect the arrangement of the intertwined knots in the two systems to become more distinguishable when the polymer stiffness is increased. As shown in Figure 5.9, this is indeed the case. Interestingly, we observe that both systems present a free energy minimum for a value of θ between 30° and 45°, indicating that their loops tend to be one inside the other; this minimum is more pronounced and corresponds to smaller angles for the $(+-)$ system, suggestive of more planar configurations. Furthermore, the minimum energy angle θ^* decreases with increasing bending energy, meaning that the aperture of the cone between the two vectors decreases with increasing polymer stiffness. The knots therefore get more aligned with increasing bending energy, and for opposite relative chiralities.

Investigating the behavior of $\theta_\perp$, we note that while the $(++)$ system presents a minimum in $\theta_\perp = 0$, the $(+-)$ system presents two minima for $\theta_\perp \sim \pm 30°$. The minima are shallow in both systems, and therefore all values of $\theta_\perp$ are easily explored, although the barriers seem to be lower in the $(+-)$ system. The presence of two minima in the $(+-)$ system can be understood considering that for it there can be two distinct arrangements of the intertwined knots: one in which the 3_1^+ is nested inside the 3_1^- knot, and the other in which the opposite happens. For every configuration with the 3_1^+ nested inside the 3_1^-, there is a corresponding specular configuration in which the handedness of the two knots is reversed and the 3_1^- is nested inside the 3_1^+. These specular configurations have the same Boltzmann weight. Given our definition of the sign of $\theta_\perp$, these two arrangements (being

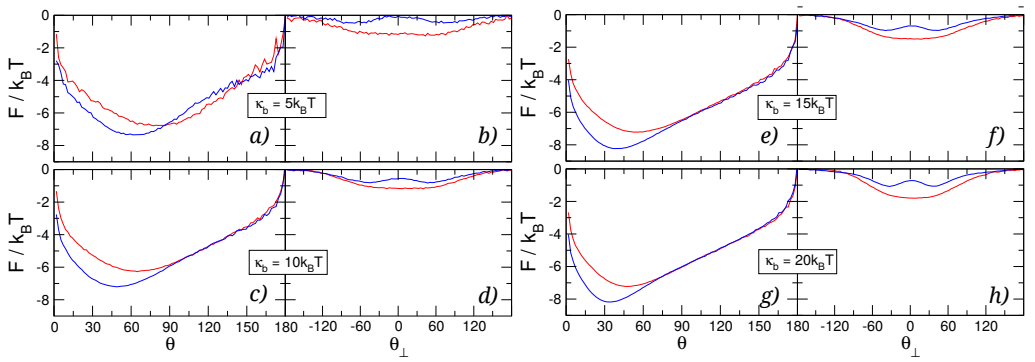


Fig. 5.9 Free energy of two intertwined trefoil knots with the same handedness (red curves) and opposite handedness (blue curves) as a function of θ (left column) and $\theta^{\perp}$ (right column), for four different values of polymer bending rigidity: from top to bottom, $\kappa_b = 5$, $\kappa_b = 10$, $\kappa_b = 15$, and $\kappa_b = 20 k_B T$. The free energies have been shifted so that their maximum values correspond to zero.

specular) will correspond to opposite values of $\theta_{\perp}$. The values of $\theta_{\perp}$ in which two minima are located are almost exactly coincident to those for which the free energy of the $(+-)$ system as a function of θ has its minima (see Figure 5.9a,c,e,g). This is suggestive of the fact that the **U** directors lie mainly in the YZ plane, so that their projections form the same angle as the three-dimensional vectors themselves. On the contrary, the broad, single minimum in $\theta_{\perp} = 0^{\circ}$ of the free energy of the $(++)$ system is due to a more skewed arrangement of the nested knot, whose director—together with that of the larger knot—identifies a plane roughly perpendicular to the YZ plane; the projections of the two directors thus form a quite small angle—zero on average—in spite of a wider angle $\theta \sim 45^{\circ}$ in the three dimensional space. This arrangement is consistent with the two intertwined prime components of the $(++)$ system to be less coplanar.

5.4 Discussion

The detailed analysis of the free energy of semiflexible chains under strong external tension as a function of various collective order parameters has led us to identify several different regimes:

- When the knots along the chain are clearly separated and sufficiently far from the hard walls, the free energy is dominated by an *entropic interaction*

between the knots, dependent on the absolute linear distance between them as well as the length of the knots.

- As the knots get closer to one another, but can still be considered as two simple separate knots, a repulsive interaction starts to dominate the free energy, stemming primarily from the steric hindrance of the proximal loops of the knots.

- Finally, as the knots become intertwined at yet smaller effective separations, the absolute magnitude of the free energy—and consequently, the stability of the knotted chain configuration—is dependent both on the bending stiffness of the polymer chain as well as on the *relative chirality* of the two knots.

In what follows, we will provide simple, mechanistic reasons for this behavior based on considerations of the various contributions of different knotted configurations to the total elastic energy of the chain.

5.4.1 Elastic Energy Model for the Size of Two Separate, Non-Interacting Knots

We start our discussion by observing that the behaviour of two separated knots along a polymer, stretched by high enough imposed forces to localize the knot, is essentially dictated by the knot length, as indicated both by Equation (5.8) and by the fact that the interaction distance scales with the knot length, as reported in Figure 5.6. It is therefore of interest to derive a simple expression for the expected knot length as a function of the bending rigidity of the chain.

The bending rigidity of a knot can be computed by modeling it as a loop of radius R, and taking into account that its length (as provided by the location algorithm) is not equal to $2\pi R$, but also includes the braided region. We approximate the extent of the braided region with a segment of length $\sim 2R$, and assume it has the same curvature as the rest of the knot. These approximations lead to a knot length $L = (2\pi + 2)R$, and the radius (expressed as a function of the length) as $R = L/(2\pi + 2)$. The bending energy then follows as:

$$E_b = \int_L dl \, \frac{1}{2}\kappa_b \left(\frac{1}{R}\right)^2 = \frac{2(\pi+1)^2 \sigma \kappa_b}{n} \tag{5.12}$$

where we used $L = \sigma n$, with n number of beads of the knot and σ the bond length.

The (effective) free energy of the knotted chain as a function of the knot length is given by the sum of the bending energy contribution of each knot, plus the contribution of the whole chain due to external tension. For the latter, we assume that a linear force acts on the knot as a function of its size, thus leading to a quadratic free energy of the form [29]:

$$E_t = \frac{1}{2} Q n^2, \tag{5.13}$$

where Q is an as-yet unspecified constant. The total free energy of two independent and equivalent knots is then given by:

$$E_{\text{tot}} = \sum_{i=1,2} \left(\frac{2(\pi+1)^2 \sigma \kappa_b}{n_i} + \frac{1}{2} Q n_i^2 \right). \tag{5.14}$$

To compute the average size of the knots, we investigate the minimum of E_{tot} with respect to n_1, n_2. We get straightforwardly that the above free energy is minimized by:

$$n_{1,2}^{\star} = \left(\frac{2(\pi+1)^2 \sigma \kappa_b}{Q} \right)^{1/3}. \tag{5.15}$$

The value of Q can then be obtained by fitting the numeric knot size; e.g., taking the value of $n \sim 36$ for $\kappa_b = 20$, see Figure 5.2a. This gives us $Q = 0.015$. The resulting function, reported in Figure 5.2a, is in very good agreement with the data obtained in the simulations.

5.4.2 Elastic Energy Model for Chirality Effects in Knot–Knot Interaction

Having analytically characterized the behaviour of two separated knots on a semi-flexible chain under external tension, we now proceed to discuss the underpinnings of the dependence of the stability of the intertwined state on the relative chirality of the two knots. While a general elucidation and the corresponding analytical

formulation of this effect will be hard to come by—being dependent on the complex details of the polymer configuration—we nevertheless offer a simple insight which to some extent clarifies the ultimate mechanical reason giving rise to this phenomenon, without delving too deeply into all the pertinent details.

We start by observing that a torus knot braid (portion of the knot containing its essential crossings) can be modeled as a helix, as can be seen from Figure 5.10. For simplicity, we envision the two trefoil knots as having their braids lying in the same plane (i.e., the axes of their helices are coplanar). We now concentrate on the portion of the chain connecting the two braids, colored in the upper part of panels (b) and (c) of Figure 5.10. Rotating the axes of the braids in such a way that they are directed into the page, we obtain the schematics of the lower part of panels (b) and (c). Obviously, the trajectory of the chain segment connecting the two is quite different for the case of opposite and equal chiralities, since the braids are rotated around their axes by an angle π or 0 for opposite and same chiralities, respectively. In order for the segment of the chain to seamlessly connect the two braids, its spatial configuration must be fundamentally different, as indicated in the bottom parts of panels (b) and (c).

The reason why two types of composite knots can have different elastic energies can then be gleaned from a highly idealized configuration, where the composite knot is presented by loops and braids (Figure 5.10). Preserving the chirality distribution in the composite knot, one can discern that the elastic energy of a simplified $(+-)$ configuration has to be smaller than for the $(++)$ configuration. Concentrating exclusively on the segment connecting the two braids in the composite knot, we can write its bending energy as:

$$E_b = \int_L dl\, \frac{1}{2}\kappa_b \left(\frac{1}{R}\right)^2 = \int_L dl\, \frac{1}{2}\kappa_b \left(\frac{d\psi(l)}{dl}\right)^2, \tag{5.16}$$

where $\psi(l)$ is now the angle between the horizontal axis and a position on the connecting segment at l. The minimization of this bending energy (leading to $d^2\psi(l)/dl^2 = 0$) should be accompanied by the appropriate boundary conditions and symmetry of the solution $\psi(l)$. This problem also bears some distant similarity to the wrapping transition and wrapping-mediated interactions for discrete binding along an elastic filament [51], where the role of the wrapping adsorbands would be played by the constituent braids of the knot.

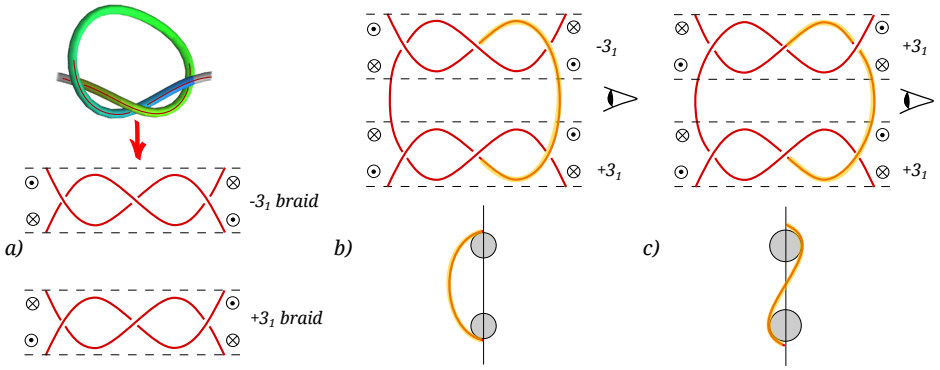


Fig. 5.10 (**a**) A trefoil knot can be decomposed into two loops and a braid. The two enantiomers are defined by the braid. The circles indicate the orthogonal components of the bending force vector. Following the usual conventions, crossed circles indicate chain directors going into the page, while dotted circles indicate chain directors coming out of the page; (**b**,**c**): Composing the intertwined knots of equal and/or different chiralities. In an intertwined $(+-)$ knot (**b**), the director joins the junction between the braid and the loop in the same sense, while in an intertwined $(++)$ knot (**c**), it points in opposite sense at the two junctions; (**b**,**c**) bottom: Schematic representation of the axial projection (in the direction of the braid axes, assumed to coincide) for the $(+-)$ and $(++)$ composite knots, respectively.

Clearly, the $(++)$ configuration corresponds to a node of a curvature, while the $(+-)$ configuration does not. In general, nodes make the energy higher, and we therefore conclude that the $(++)$ configuration would have a higher bending energy. One must anyway keep in mind that this argument is only approximate and would need to be refined, since we are not allowing the braids to relax in order to minimize the bending energy. Admittedly, then, the above argument is very simplistic, but sufficiently robust to explain the observed knot interaction energies obtained with a full MD tracing of the partition function of the knots.

5.5 Conclusions

By performing detailed underdamped MD simulations of a constrained semiflexible chain under strong external tension, we investigated its free energy as a function of a properly-defined collective order parameter, $|D|$, quantifying the absolute linear distance between the centers of two knots along the chain. We discovered several regimes in this dependence, characterized by the relative importance of entropy,

bending energy, and end constraints. In each of these regimes, the free energy depends in a characteristic way on $|D|$.

The first regime is characterized by knots being clearly separated along the chain; i.e., no intertwining, with the free energy being dominated by a non-specific, entropic interaction between them as well as by the end constraints when the knots are close to the confining walls. When the separation between knots diminishes, a repulsive interaction emerges due to the steric hindrance between close loops of the knots. As the knots start to intertwine at smaller effective separations, the free energy is dominated by the bending energy, dependent both on the bending stiffness of the polymer chain as well as on the relative chirality of the two knots.

Emergence of chirality as the defining factor for the stability of intertwined knots is the most important corollary of our work. In this sense, the knots behave in a fundamentally different way from sliplinks [133], obviously bearing another degree of freedom that affects their interaction in a fundamental manner.

We analyzed the intertwined state by introducing two further order parameters, θ and $\theta_{\perp}$, which capture the relative orientation between the two knots. Mapping the free energy of the system onto them, we observed that two intertwined trefoils tend to have their loops one inside the other, and that this tendency increases with increasing bending energy. Furthermore, two trefoil knots of opposite chirality tend to be in a more planar configuration than two trefoils having the same chirality.

Finally, we identified the mechanistic basis for the chiral effect in the intertwined state by observing that the braids of two torus knots of opposite chirality have different alignment with respect to the tensioned rope on which they are tied. Because of this, two knots of opposite chirality can form an intertwined state in which their braids are aligned without introducing any additional bending in the chain, since they are the specular image of one another. The same cannot happen for two knots of the same chirality, which can only be connected with their braids lying on the same plane by introducing a node in the curvature of the segments connecting the braids, thus increasing the bending energy.

5.6 Acknowledgments

Luca Tubiana and Rudolf Podgornik acknowledge support from the Slovenian Agency for Research and Development (ARRS grant No. J1-4134). Luca Tubiana acknowledges also support from the Mahlke-Oberman Stiftung and the European Union's Seventh Framework Programme for research, technological development and demonstration (grant No. 609431). Saeed Najafi and Raffaello Potestio are thankful to Roberto Menichetti for an attentive reading of the manuscript and useful comments.

Chapter 6

Entanglement of knotted DNA ring and an entwined DNA loop

To be submitted as:
Saeed Najafi and Raffaello Potestio
Entanglement of knotted DNA ring and an entwined DNA loop

Abstract

By means of computer simulations, in this study we demonstrate that the crossing pattern of a braid that consists of entwined DNA rings play a major role in its structural and dynamical properties. In particular, we found that the braid with non-identical crossing pattern, that the latter is specified by comparing the partitioned inner and outer cross sections, enforces a positive and stronger correlation between the entangled rings.

6.1 Introduction

Very recently, many scientists have been attracted to the mystery of knots and also challenged by its difficulties. The entity of topology and the related geometry, is in fact a matter of interest in different fields of Soft Materials from liquid crystals [165, 87], up to bio-polymers such as proteins, DNA and RNA filaments [157, 33, 40, 147, 160, 144, 110, 106, 105]. The conjunction of topologies and entanglement

in bio-polymers are indeed ubiquitous in biological processes [33, 34, 149, 105, 52, 159, 46]. During Meiosis, genetic informations on the chromosomes are believed to be inherited in cross-over process that is known as Prophase I; in such a genetic exchange between DNAs, the impact of the entanglement pattern and subsequently the effect of induced strain on the entangled chromosomes is largely unknown [8, 6].

Geometrical and dynamical properties of DNA are presumably important during replication, transcription and in the other aspects of its metabolism [101, 80, 31]. DNA can make unknotted loops and regulate its transcription [101], however, it can also be knotted [33, 149, 107, 123, 152]. Braids of entangled knot with unknotted DNA loop can be formed in DNA transactions [71], as well as in bacteriophages [99, 27, 28]. Depending on the interplay between entwined topological defects in the braid, the entangled DNAs can exhibit a variety of conformations which permit them to carry out specific functions. Understanding the relation that links the topological constrains to their structural and dynamical features is a difficult task that can be investigated by means of computer simulations, where various conformational features of topologies can be measured and related to structure and dynamic emergent properties.

In this article, we attempt to study the impact of minimal crossing pattern of braids of DNAs, on their structural and dynamical features. The braid consists of a knot and a pierced unknotted DNA through the all polygons of the knot and then closed, see Fig. 6.1. The corresponding crossing pattern of the braid can be drawn by making use of the Right-hand rule as depicted in the top of Fig. 6.1. We use the common bead-spring model of polymer chains that is introduced by Kremer and Grest [65]. Additionally, we exert a bending potential to the chains and exam our systems at different salt concentrations (at extremes: in poor solvent with no salt $\kappa = 0$ and in good solvent at high salt $\kappa = 20k_BT$, where κ is the bending stiffness). The length of the knot in all setups is 70σ and the size of the entwined loop is 30σ, where σ is the unit of the length.

In Fig 6.1, we illustrate a view of the systems that are under study. In panels a, b, c and d the 3_1^+, 3_1^-, 5_1^- and 5_1^+ knots with the entwined loop are shown, respectively. The minimal crossing pattern of the braids is mapped onto the top left of each panel. In the case of trefoil (resp-5_1) knot, the entwined loop has a

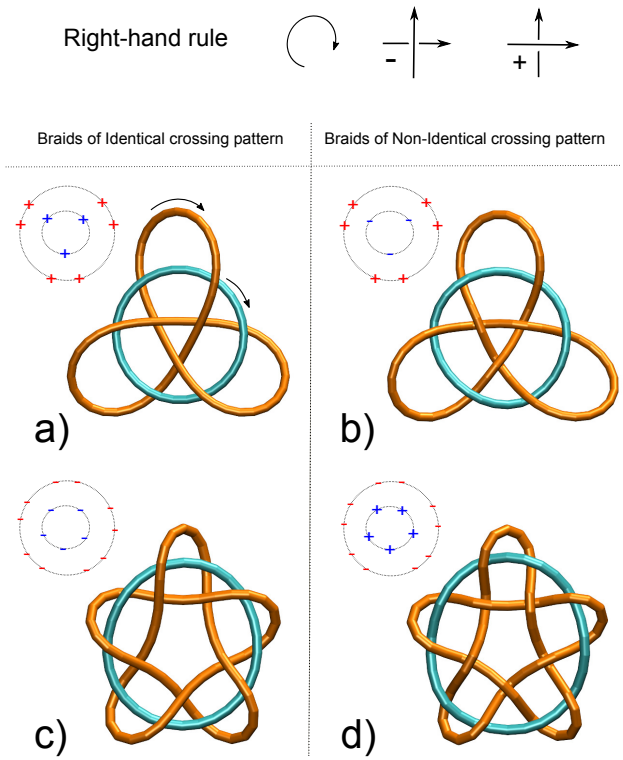


Fig. 6.1 Snapshots of the systems under examination. In panel a, b, c and d we show the 3_1^+, 3_1^-, 5_1^- and 5_1^+ knots and the entwined loop, respectively. In the top left of each panel we projected the minimal crossing pattern of the braid by using the Right-hand rule that is depicted in the top of the figure. The length of the knot in all setups is 70σ and the size of the loop is 30σ.

positive (resp-negative) crossing pattern, however, switching the knot chirality will establish identical and non-identical crossing pattern for the braids.

6.2 Results and discussion

Energy Analysis $\rightarrow$ Topoisomerase proteins are responsible for DNA unknotting. This has been speculated that topoisomerase has higher affinity to prebent domains of DNA, meaning that knotted portion of DNA that is tightly bent is very likely to be uncrossed by topoisomerase. Considering this relevance, we start our study by probing the bending energy difference between the braid that consists of a knotted DNA and an entwined DNA loop and the same system when the rings are separated ($\delta U_{bend} = U_{bend}^{entwine} - U_{bend}^{sep}$). In Fig 6.2 plot a and b, we show the latter quantity at different salt concentration of the systems that contains knot 3_1 and knot 5_1,

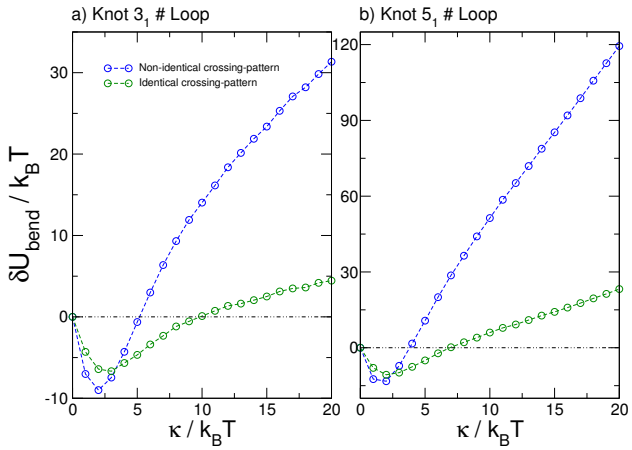


Fig. 6.2 Plot a (resp-b), shows the bending energy difference between the braid of knot 3_1 (resp-5_1) and the corresponded separated rings. The green and blue curves indicates the bending energy difference of the braids with identical and non-identical crossing pattern.

respectively. In both braids with identical and non-identical crossing pattern, a phase transition occurs from energetically favored entangled conformation to a tense phase that in which the bending energy of the entangled rings is higher than the separated state. Specially, the braids of knot 5_1 due to the complexity of the knot are remarkably tighter and tenser than the braids of knot 3_1 which indicates an overall enhancement of the unknotting propensity in presence of topoisomerase.

The rationale behind the phase transition in the conformational energy as function of salt concentration of the system, can be understood in terms of the interplay between the conjuncted topologies that enforces the structure of the braids. In fact at the low bending stiffness, the entangled knot and loop can give structure to the braids that will prevent the steric interactions between the different segments of entwined DNAs and leads to a decrease in the conformational energy. The presence of the minimum reveals that the braids within given topologies can adequately adjust their conformational energy with the related geometry and thereafter, as the consequence of specific entanglement in the braids with non-identical crossing pattern, the incurred bending energy cost of the entwinement grows rapidly. Here one can conclude that depending on salt concentration of the system the crossing pattern of entwined DNAs might either attenuate or facilitate the uncrossing process by topoisomerase; we emphasize that the previous implications of conjuncted topologies are doomed to be less evident in the braids with identical crossing pattern.

Dynamics and Structure $\rightarrow$ We quantify and capture the dynamics between the entangled rings by studying the correlation coefficient of collective variables of the system that are considered to be the radius of gyration of the knot (KR_g) and the radius of gyration of the loop (LR_g). In Fig. 6.3, panel I (resp-II), plot a, we show the correlation coefficient between knot 3_1 (resp-5_1) and the entwined loop in different bending stiffness of the chains by making use of Eq. (6.1):

$$\mathrm{Corr}(\mathrm{KR_g}, \mathrm{LR_g}) = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \Delta_{\mathrm{i}}(\mathrm{KR_g}).\Delta_{\mathrm{i}}(\mathrm{LR_g})}{\sqrt{(\sum_{\mathrm{i}=1}^{\mathrm{n}} [\Delta_{\mathrm{i}}(\mathrm{KR_g})]^2).(\sum_{\mathrm{i}=1}^{\mathrm{n}} [\Delta_{\mathrm{i}}(\mathrm{LR_g})]^2)}} \tag{6.1}$$

where Δ_i refers to the difference between the ith sample and the average value.

As a remarkable result, the correlation coefficient in the braids with non-identical crossing pattern are decidedly strong and positive, which a posteriori enforces tuneful intra-dynamics between the knots and the entwined loop. Hence, one can perceive that during Prophase I process, the non-identical crossing pattern of entangled chromosomes might engender appropriate constrains for effective genetic exchange between DNAs. In contrast, in the braids with identical crossing pattern, we observe a purely negative correlation coefficient between knot 3_1 and the entwined loop, particularly, in the braid of knot 5_1 the correlation coefficient turns to positive values at $\sim 14 k_B T$.

We provide a simple qualitative and intuitive explanation for the negative and positive correlation coefficients of the dynamics between knot 3_1 and the entwined loop. The topological constraints on the braids compels spontaneous negative and positive curvatures in the linkage of the crosses. Specifically, those under-over crosses of the loop-knot and knot-knot crosslet, where are in taken by the concave of the mediated linkage (shown in the top of Fig. 6.4), effectively will enforce a geometrical entanglement between the knot and the loop. The loop-knot and knot-knot crosslet that enforces the effective entanglement between the knot and the loop is indicated in the top of Fig. 6.4. The blue and red circles in the crosslet shows the cross of the loop and knot, and the knot with itself, respectively, the crosses of the crosslet that makes the effective entanglement must be under and over (or opposite)

in respect to the mediated linkage, otherwise due to the geometry of the mediated linkage which is enforced by the conjuncted topologies, there wouldn't be any effective entanglement between the crosses of the knot and loop in the braid. For clarifying the appearance, this crosslet is encoded by a simple spring in the braids. The red and blue sketches in bottom panel of Fig. 6.4 shows a surmised elastic rod model for knot 3_1 and the loop, respectively. In panel a, in the braid with identical crossing pattern one can clearly see that for instance by dragging the three corners of the red shape that is equivalent to increasing the radius of gyration of the knot, due to the specific entanglement the radius of gyration of the entwined loop (blue dashed line) decreases, that entailing to a negative correlation coefficient. However in panel b, in the braid with non-identical crossing pattern, increasing the radius of gyration of the knot will lead to an increase in the radius of gyration of the loop as well, that causes a positive correlation coefficient.

The non-monotonic behavior in the correlation profile can be understood by investigating the relaxation length of the mediated springs that is essentially the effective length of the linkage of the under-over crosses in the loop-knot and knot-knot crosslet, which as a geometric feature depends on the relative length of the knot and the entwined loop. In the current setups, in the braid with identical crossing pattern, one can see that the minimum of the correlation coefficient between knot 3_1 and the loop located at $\sim 5-6\frac{l_p}{\sigma}$, where l_p is the persistent length of the chain. However, in the braid with non-identical crossing pattern, due to the closeness of under-over crosses in the loop-knot and knot-knot crosslet, reasonably, the length of the related linkage should be shorter, that the predicted quantity is confirmed by the position of the minimum correlation coefficient where is located at $\sim 3-4\frac{l_p}{\sigma}$. Basically, same justification can be applied in the braids that contains knot 5_1, however, because of the complexity of knot 5_1 the finite size effect will play a crucial role.

The dynamical properties of the entwined rings provides specific structural features for the braids. Here the latter is probed by characterizing the harmonious modulation of the chains that is considered as the relative strain of the knot and the entwined loop: $\xi = \left(\frac{stdev.KR_g}{<KR_g>}\right) \cdot \left(\frac{stdev.LR_g}{<LR_g>}\right)^{-1}$. In Fig. 6.3, panel I (resp-II), plot b, we show the relative strain of knot 3_1 (resp-5_1) and the entwined loop. Consistent with the correlation profile, in the braid with identical crossing pattern by increasing the bending stiffness of the chains the difference in the strain of the knot and the loop increases. However, in the braid with non-identical crossing

pattern due to the strong correlation in the conformational changes, the relative strain has this affinity to be close to 1.

Further insight comes from the analysis of the knots configuration in the braids. In Fig. 6.3, panel I (resp-II), plot c and d, we investigate the free energy of the knot at $\kappa = 20k_BT$ in the braids with identical and non-identical crossing pattern, respectively, by considering the radius of gyration of knot 3_1 (resp-5_1) and the absolute distance between center mass of the knot and the entwined loop as the collective variables. In both braids, we distinctly observed a favored constricted conformation of the knot, particularly in the braid with non-identical crossing pattern the knot is strongly shrunk. As discussed earlier, a possible explanation for the tendency of the knots toward the constricted conformation in the braids of non-identical pattern is that the bending energy of the latter is much higher than the other system, which enforces a narrow range for knots conformational fluctuations that would leads to a sharper funnel of the free energy.

Discussion $\rightarrow$ In addition to the braids that consists of a knot and an entwined loop, we would like to study other entangled chains and try to corroborate our claim regarding the implications of crossing pattern. Intending to this aim, we consider two systems of three linked rings with 6 minimal crossings, where the size of each ring is 30σ. The first setup is three entangled rings that its link number is $+3$ (Hopf-linked rings), however, the link number of the next system is 0 (Borromean rings). In Fig. 6.5, we report the radius of gyration of the Hopf-linked rings and the Borromean rings, versus to bending stiffness of the chain, green and blue curves, respectively. As it is clear from plot, the Hopf-linked rings with identical crossing pattern shows a higher radius of gyration, in compare to the Borromean rings with non-identical crossing pattern. The Borromean system is strongly constricted and does a well formed structure at high bending stiffness that is known as *Monkey Fist*. The fluctuations in the radius of gyration of the Borromean rings decreases by increasing the bending stiffness of the chain, although the Hopf-linked rings are highly fluctuative. The Borromean rings due to their non-identical crossing pattern are strongly entangled, that the impact of this effect manifests itself mostly in the pairwise distance between the entwined rings. As one can see from inset plot of Fig. 6.5, the difference between the radius of gyration of a single ring in the Hopf-linked and the Borromean systems is negligible, however, there is a drastic difference in the radius of gyrations of the total linked rings.

6.3 Concluding remarks

In conclusion, we found that by designing a specific crossing pattern the dynamics of entwined polymers can subtly be modulated, that the latter essentially could derive the entangled system to the certain structural properties. Since such a conformational features appears to be crucial in biological systems, this investigation could be fruitful in understanding the role of entanglement in the function of entwined biopolymers.

6.4 Materials and Methods

6.4.1 Model and simulation details

We use Kremer-Grest model for our polymer chains [65], that enforces the consecutive bounds between the beads by *fene* potential and the chains prevented to be self crossed by employing repulsive Weeks-Chandlers-Anderson (wca) potential. The total potential energy of the chains is:

$$\mathscr{H} = U_{wca} + U_{fene} + U_{bend} \tag{6.2}$$

The wca potential is given by:

$$U_{wca} = \frac{1}{2} \sum_{(i,j),j\neq i}^{N} V(d_{i,j}), \tag{6.3}$$

$$V(r) = \begin{cases} 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + \frac{1}{4} \right] & \text{for } r \leq 2^{1/6}\sigma \\ 0 \ \text{ otherwise} \end{cases} \tag{6.4}$$

where $\varepsilon = 1$ sets the energy scale. The *fene* potential reads:

$$U_{fene} = -\sum_{i=1}^{N-1} \frac{\kappa_{fene}}{2} \left(\frac{R_0}{\sigma}\right)^2 \ln \left[1 - \left(\frac{d_{i,i+1}}{R_0} \right)^2 \right] \tag{6.5}$$

where $d_i = |\vec{r}_i - \vec{r}_{i+1}|$ is the distance of the bead centers i and $i+1$, $R_0 = 1.5\sigma$ is the maximum bond length and $\kappa_{fene} = 30\varepsilon$ is the interaction strength.

The bending rigidity potential is defined as:

$$U_{bend} = \sum_{i=1}^{N-1} \kappa \left(1 - \frac{\vec{d}_i \cdot \vec{d}_{i+1}}{|\vec{d}_i||\vec{d}_{i+1}|} \right) \tag{6.6}$$

where κ is the bending stiffness of the chain.

The MD simulations are carried out with Lammps package integrating the Langevin equations of motion in constant temperature with $k_B T = \varepsilon$ and $\tau = \sigma\sqrt{m/\varepsilon} = 1$ MD time units.

For each setup in certain bending stiffness, we ran 5 independent simulations that each covers 2.10^9 time steps, an equilibration short simulation that covers $10^6\tau$ was also performed. The applied time step is $\Delta t = 0.01\tau$, and the analyzes are carried out at every 100τ.

6.4.2 Computation of Free energy, Radius of Gyration and Center of Mass

In order to characterize the free energy of the system, we have computed the quantity that defined as:

$$F(x,y) = -k_B T \ln P(x,y) \tag{6.7}$$

$P(x,y)$ is the instantaneous probability distribution of collective variables that are system specific.

The radius of gyration and center mass of the rings is computed as follow:

$$R_g = \sqrt{\left\langle \frac{1}{2M^2} \sum_{i=0}^{M} \sum_{j=0}^{M} |\vec{r}_i - \vec{r}_j|^2 \right\rangle} \tag{6.8}$$

$$CM = \frac{1}{M} \sum_{i=0}^{M} \vec{r}_i \tag{6.9}$$

where i and j are the index of the beads and M is the number of beads that is system specific.

6.5 Acknowledgments

We are grateful to Aoife Fogarty for critically reading the manuscript of the paper.

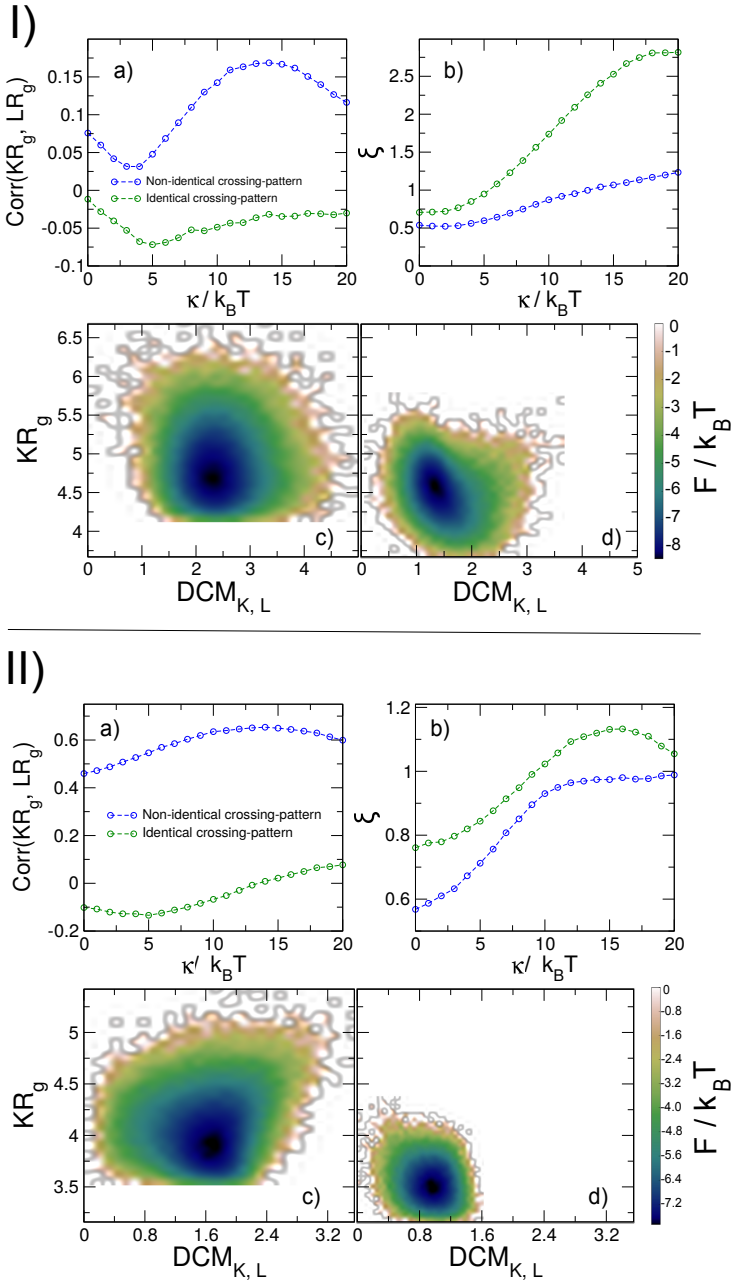


Fig. 6.3 In panel I (resp-II), plot a and b shows the correlation coefficient and the relative strain of the 3_1 (resp-5_1) knot and the entwined loop, versus to bending stiffness of the chains, respectively. In plot c and d, we show the free energy of the braids with identical and the non-identical crossing pattern at $\kappa = 20k_BT$, respectively. The latter is carried out by studying the radius of gyration of the knot (KR_g) and the absolute distance between center mass of the knot and the loop $(DCM_{K,L})$.

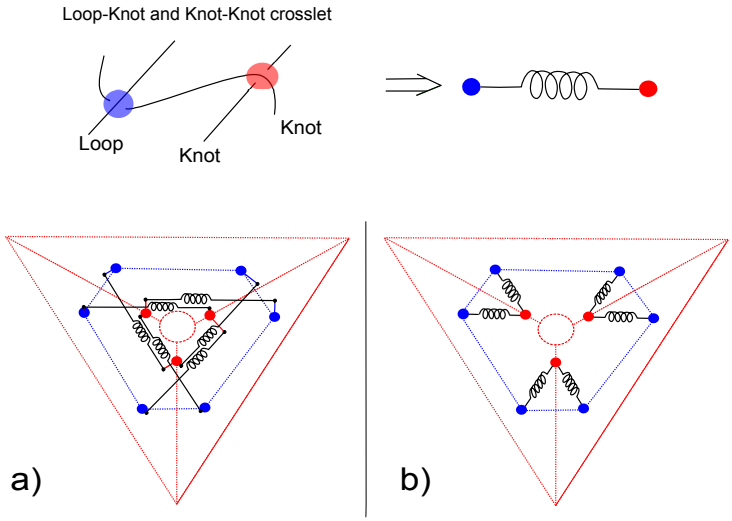


Fig. 6.4 Red and blue dashed lines shows a surmised elastic rod model for knot 3_1 and the loop, respectively. Red and blue circles shows the cross between the knot and itself, and the loop with knot, respectively. The geometrical entanglement between knot 3_1 and the entwined ring is encoded by the simple springs where are depicted between the in taken crosses by the concave of the mediated linkage in the braids with identical and non-identical crossing pattern, panel a and b, respectively.

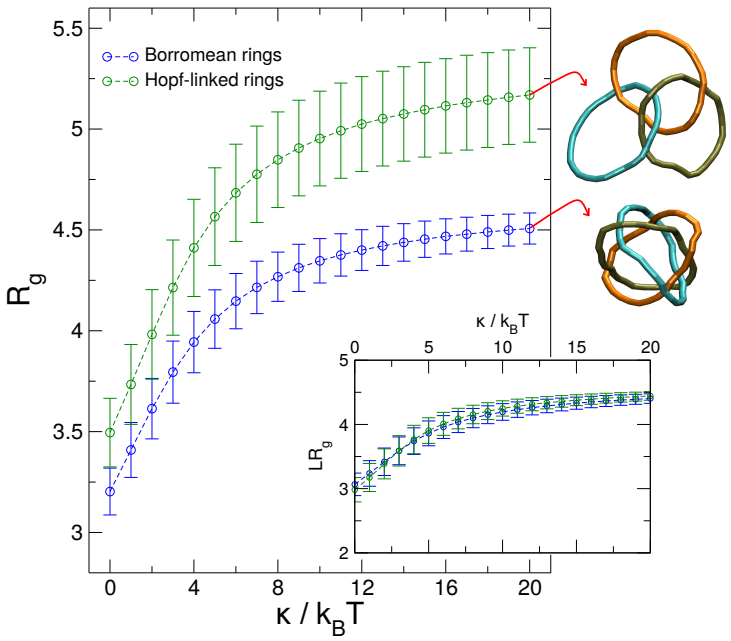


Fig. 6.5 Here we show the radius of gyration of three Hopf-linked rings with identical crossing pattern and the radius of gyration of the Borromean rings with non-identical crossing pattern versus to bending stiffness of the chains, green and blue curves, respectively.

Chapter 7

Conclusion

Topological constraints, which often accompany biological activity, are at the core of our study. The characterization of the structural and dynamical properties of topological constraints can provide a comprehensive insight into their impact on the function and duties of entangled macromolecules. In this thesis, we presented numerical investigations to identify the key features of self-entangled and linked bio-polymers in a variety of biological systems.

The sequence of a polymer not only determines its geometrical structure but also its topology. In particular, understanding the folding of a polypeptide chain into knotted conformation driven by relatively simple native and non-native interactions is a major endeavor that demands an accurate explanation. We introduced a novel protocol to encode global structure of knotted proteins as a set of local characteristic information of the chain. Consistent with realistic all atom simulations, our approach was capable of reproducing the optimal folding pathway of two short knotted proteins.

We aimed to reduce the complexity of the knotting process in dsDNA by introducing the smallest amount of sequence information along the chain. The rich variety of the knot spectrum featured by simple DNA with site-specific interactions reveals a remarkable insight to build a basic understanding of the relation between the building blocks of a macromolecule and its topology.

Emergence of entanglement pattern as the control factor for the dynamical and structural properties of topologically constrained biopolymers is an important result of our study. In particular, we found that depending on the relative chirality

of knots on a stretched chain, the knots can interact in a fundamentally different way. Finally, to rationalize the chirality effect in the intertwined state of the knots, we studied the braids of entwined knotted and unknotted DNA rings with identical and non-identical crossing pattern.

Our result motivate further efforts, especially for experimentalist to elucidate bimolecular knotting and also make further advancements toward designing topologically constrained macromolecules with specific function in biological systems.

References

[1] Rubinstein, M. and Colby, R.H. (2003). Polymer Physics. *OUP Oxford.*

[2] Doi, M. and Edwards, S.F. (1988). The Theory of Polymer Dynamics. *Clarendon Press.*

[3] Clayden, J. and Greeves, N. and Warren, S. (2012). Organic Chemistry. *OUP Oxford.*

[4] Flory, P. J. and Vrij, A. (1963). Melting Points of Linear-Chain Homologs. The Normal Paraffin Hydrocarbons. *Journal of the American Chemical Society*, 85(22):3548–3553.

[5] Buchmeiser, M.R., Baughman, T.W., Buchmeiser, M.R., Fuchs, G., Riegler, S., Slugovc, C., Stelzer, F., Trimmel, G. and Wagener, K.B. (2005). Metathesis Polymerization. *Springer Berlin Heidelberg.*

[6] Stephen R. Bolsover, Jeremy S. Hyams, Elizabeth A. Shephard, Hugh A. White, Claudia G. Wiedemann (2004). Cell Biology: A Short Course. *John Wiley & Sons, Inc*.

[7] Lodish, H. and Berk, A. and Kaiser, C.A. and Krieger, M. and Scott, M.P. and Bretscher, A. and Ploegh, H. and Matsudaira, P. (2013). Molecular Cell Biology. *W. H. Freeman*.

[8] Bruce Alberts, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, Peter Walter (2014). Molecular Biology of the Cell. *Garland Science*.

[9] Chothia, Cyrus and Gough, Julian and Vogel, Christine and Teichmann, Sarah A. (2003). Evolution of the Protein Repertoire. *Science*, 300(5626):1701–1703.

[10] Rubinow, S. (1994). The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots. *W.H. Freeman.*

[11] Cromwell, P. R. (2004). Knots and Links. *Cambridge University Press.*

[12] Livingston, C. (1993). Knot Theory. *Mathematical Association of America.*

[13] J. W. Alexander (1928). Topological Invariants of Knots and Links. *Transactions of the American Mathematical Society.*, 30(2):275–306.

[14] De Witt Sumners (1995). Lifting the curtain: using topology to probe the hidden action of enzymes. *Notices of the AMS*, 42:528–537.

[15] Huang, K. (1987). Statistical mechanics. *Wiley.*

[16] Christine A. Orengo and Janet M. Thornton (2003). PROTEIN FAMILIES AND THEIR EVOLUTION–A STRUCTURAL PERSPECTIVE. *Annual Review of Biochemistry*, 74(1):867–900.

[17] Antonina Andreeva and Alexey G Murzin (2006). Evolution of protein fold in the presence of functional constraints. *Current Opinion in Structural Biology*, 16(3):399–408.

[18] Sophie E Jackson, Antonio Suma, and Cristian Micheletti. (2017). How to fold intricately: using theory and experiments to unravel the properties of knotted proteins *Current Opinion in Structural Biology*, 42(1):6-14.

[19] Banavar, Jayanth R. and Maritan, Amos and Micheletti, Cristian and Trovato, Antonio. (2002). Geometry and physics of proteins. *Proteins: Structure, Function, and Bioinformatics*, 47(3):315–322.

[20] F. Seno and A. Trovato. (2007). Minireview: The compact phase in polymers and proteins. *Physica A: Statistical Mechanics and its Applications*, 384(1):122–127.

[21] Francesco Pontiggia and Andrea Zen and Cristian Micheletti. (2008). Small- and Large-Scale Conformational Changes of Adenylate Kinase: A Molecular Dynamics Study of the Subdomain Motion and Mechanics. *Biophysical Journal* , 95(12):5901–5912.

[22] M Delarue and Y.-H Sanejouand. (2002). Simplified Normal Mode Analysis of Conformational Transitions in DNA-dependent Polymerases: the Elastic Network Model. *Journal of Molecular Biology*, 320(5):1011–1024.

[23] Javier Arsuaga, Robert K.-Z Tan, Mariel Vazquez, De Witt Sumners and Stephen C Harvey. (2002). Investigation of viral {DNA} packaging using molecular mechanics models. *Biophysical Chemistry*, 101-102:475–484.

[24] Rolf Auhl, Ralf Everaers, Gary S. Grest, Kurt Kremer and Steven J. Plimpton. (2003). Equilibration of long chain polymer melts in computer simulations. *The Journal of Chemical Physics*, 119(24):12718–12728.

[25] a Beccara, S., Škrbić, T., Covino, R., Micheletti, C., and Faccioli, P. (2013). Folding pathways of a knotted protein with a realistic atomistic force field. *PLOS Computational Biology*, 9(3):1–9.

[26] Arai, Y., Yasuda, R., Akashi, K.-i., Harada, Y., Miyata, H., Kinosita, K., and Itoh, H. (1999). Tying a molecular knot with optical tweezers. *Nature*, 399(6735):446–448.

[27] Arsuaga, J., Vazquez, M., McGuirk, P., Trigueros, S., Sumners, D. W., and Roca, J. (2005). Dna knots reveal a chiral organization of dna in phage capsids. *Proceedings of the National Academy of Sciences of the United States of America*, 102(26):9165–9169.

[28] Arsuaga, J., Vázquez, M., Trigueros, S., Sumners, D. W., and Roca, J. (2002). Knotting probability of dna molecules confined in restricted volumes: Dna knotting in phage capsids. *Proceedings of the National Academy of Sciences*, 99(8):5373–5377.

[29] Audoly, B., Clauvelin, N., and Neukirch, S. (2007). Elastic knots. *Phys. Rev. Lett.*, 99:164301.

[30] Baiesi, M., Orlandini, E., and Stella, A. L. (2010). The entropic cost to tie a knot. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(06):P06012.

[31] Baiesi, M., Orlandini, E., and Stella, A. L. (2014). Knotted globular ring polymers: How topology affects statistics and thermodynamics. *Macromolecules*, 47(23):8466–8476.

[32] Bao, X. R., Lee, H. J., and Quake, S. R. (2003). Behavior of complex knots in single dna molecules. *Phys. Rev. Lett.*, 91:265506.

[33] Bates, A. D. and Maxwell, A. (2005). *DNA topology*. Oxford University Press, USA.

[34] Belmonte, A. (2007). The tangled web of self-tying knots. *Proceedings of the National Academy of Sciences*, 104(44):17243–17244.

[35] Ben-Naim, E., Daya, Z. A., Vorobieff, P., and Ecke, R. E. (2001). Knots and random walks in vibrated granular chains. *Phys. Rev. Lett.*, 86:1414–1417.

[36] Best, R. B., Hummer, G., and Eaton, W. A. (2013). Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of the National Academy of Sciences*, 110(44):17874–17879.

[37] Brown, P. and Cozzarelli, N. (1979). A sign inversion mechanism for enzymatic supercoiling of dna. *Science*, 206(4422):1081–1083.

[38] Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A., and Felsenfeld, G. (2002). The insulation of genes from external enhancers and silencing chromatin. *Proceedings of the National Academy of Sciences*, 99(suppl 4):16433–16437.

[39] Bustamante, C., Smith, S. B., Liphardt, J., and Smith, D. (2000). Single-molecule studies of {DNA} mechanics. *Current Opinion in Structural Biology*, 10(3):279 – 285.

[40] Bölinger, D., Sułkowska, J. I., Hsu, H.-P., Mirny, L. A., Kardar, M., Onuchic, J. N., and Virnau, P. (2010). A stevedore's protein knot. *PLOS Computational Biology*, 6(4):1–6.

[41] Caraglio, M., Micheletti, C., and Orlandini, E. (2015). Stretching response of knotted and unknotted polymer chains. *Phys. Rev. Lett.*, 115:188301.

[42] Castro, C. E., Kilchherr, F., Kim, D.-N., Shiao, E. L., Wauer, T., Wortmann, P., Bathe, M., and Dietz, H. (2011). A primer to scaffolded dna origami. *Nat Meth*, 8(3):221–229.

[43] Champoux, J. J. (2001). Dna topoisomerases: Structure, function, and mechanism. *Annual Review of Biochemistry*, 70(1):369–413. PMID: 11395412.

[44] Chothia, C. and Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *The EMBO Journal*, 5(4):823–826.

[45] Coluzza, I., van Oostrum, P. D. J., Capone, B., Reimhult, E., and Dellago, C. (2013a). Design and folding of colloidal patchy polymers. *Soft Matter*, 9:938–944.

[46] Coluzza, I., van Oostrum, P. D. J., Capone, B., Reimhult, E., and Dellago, C. (2013b). Sequence controlled self-knotting colloidal patchy polymers. *Phys. Rev. Lett.*, 110:075501.

[47] Covino, R., Škrbić, T., Beccara, S. a., Faccioli, P., and Micheletti, C. (2014). The role of non-native interactions in the folding of knotted proteins: Insights from molecular dynamics simulations. *Biomolecules*, 4(1):1–19.

[48] Dai, L., Renner, C. B., and Doyle, P. S. (2014). Metastable tight knots in semiflexible chains. *Macromolecules*, 47(17):6135–6140.

[49] Dai, L., Renner, C. B., and Doyle, P. S. (2015). Origin of metastable knots in single flexible chains. *Phys. Rev. Lett.*, 114:037801.

[50] de Villiers, E.-M., Fauquet, C., Broker, T. R., Bernard, H.-U., and zur Hausen, H. (2004). Classification of papillomaviruses. *Virology*, 324(1):17 – 27.

[51] Dean, D. S., Hammant, T. C., Horgan, R. R., Naji, A., and Podgornik, R. (2012). Wrapping transition and wrapping-mediated interactions for discrete binding along an elastic filament: An exact solution. *The Journal of Chemical Physics*, 137(14):144904.

[52] Delbrück, M. (1962). *in Mathematical problems in biological sciences*, volume 14. Proceedings of Symposia in Applied Mathematics.

[53] des Cloizeaux, J. (1981). Ring polymers in solution : topological effects. *J. Physique Lett*, 42(19):433 – 436.

[54] Di Stefano, M., Tubiana, L., Di Ventra, M., and Micheletti, C. (2014). Driving knots on dna with ac/dc electric fields: topological friction and memory effects. *Soft Matter*, 10:6491–6498.

[55] DIAO, Y., PIPPENGER, N., and SUMNERS, D. W. (1994). On random knots. *Journal of Knot Theory and Its Ramifications*, 03(03):419–429.

[56] Dommersnes, P. G., Kantor, Y., and Kardar, M. (2002). Knots in charged polymers. *Phys. Rev. E*, 66:031802.

[57] D'Adamo, G. and Micheletti, C. (2015). Molecular crowding increases knots abundance in linear polymers. *Macromolecules*, 48(17):6337–6346.

[58] Ercolini, E., Valle, F., Adamcik, J., Witz, G., Metzler, R., De Los Rios, P., Roca, J., and Dietler, G. (2007). Fractal dimension and localization of dna knots. *Phys. Rev. Lett.*, 98:058102.

[59] Farago, O., Kantor, Y., and Kardar, M. (2002). Pulling knotted polymers. *EPL (Europhysics Letters)*, 60(1):53.

[60] Faísca, P. F. (2015). Knotted proteins: A tangled tale of structural biology. *Computational and Structural Biotechnology Journal*, 13:459 – 468.

[61] Faísca, P. F. N., Nunes, A., Travasso, R. D., and Shakhnovich, E. I. (2010a). Non-native interactions play an effective role in protein folding dynamics. *Protein Science*, 19(11):2196–2209.

[62] Faísca, P. F. N., Travasso, R. D. M., Charters, T., Nunes, A., and Cieplak, M. (2010b). The folding of knotted proteins: insights from lattice simulations. *Physical Biology*, 7(1):016009.

[63] Frisch, H. L. and Wasserman, E. (1961). Chemical topology1. *Journal of the American Chemical Society*, 83(18):3789–3795.

[64] Go, N. and Taketomi, H. (1978). Respective roles of short- and long-range interactions in protein folding. *Proceedings of the National Academy of Sciences*, 75(2):559–563.

[65] Grest, G. S. and Kremer, K. (1986). Molecular dynamics simulation for polymers in the presence of a heat bath. *Phys. Rev. A*, 33:3628–3631.

[66] Grosberg, A. Y. and Rabin, Y. (2007). Metastable tight knots in a wormlike polymer. *Phys. Rev. Lett.*, 99:217801.

[67] Holwerda, S. J. B. and de Laat, W. (2013). Ctcf: the protein, the binding partners, the binding sites and their chromatin loops. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 368(1620).

[68] Hoste, J. and Thistlethwaite, M. (1999). Knotfind.

[69] Hou, C., Zhao, H., Tanimoto, K., and Dean, A. (2008). Ctcf-dependent enhancer-blocking by alternative chromatin loop formation. *Proceedings of the National Academy of Sciences*, 105(51):20398–20403.

[70] Huang, L. and Makarov, D. E. (2007). Langevin dynamics simulations of the diffusion of molecular knots in tensioned polymer chains. *The Journal of Physical Chemistry A*, 111(41):10338–10344. PMID: 17637045.

[71] Imakaev, M. V., Tchourine, K. M., Nechaev, S. K., and Mirny, L. A. (2015). Effects of topological constraints on globular polymers. *Soft Matter*, 11:665–671.

[72] Jamroz, M., Niemyska, W., Rawdon, E. J., Stasiak, A., Millett, K. C., Sułkowski, P., and Sulkowska, J. I. (2015). Knotprot: a database of proteins with knots and slipknots. *Nucleic Acids Research*, 43(D1):D306.

[73] Jawed, M. K., Dieleman, P., Audoly, B., and Reis, P. M. (2015). Untangling the mechanics and topology in the frictional response of long overhand elastic knots. *Phys. Rev. Lett.*, 115:118302.

[74] Kauffman, L. H. (1990). An invariant of regular isotopy. *Trans. Amer. Math. Soc.*, 318(417):417–471.

[75] Khatib, F., Weirauch, M. T., and Rohl, C. A. (2006). Rapid knot detection and application to protein structure prediction. *Bioinformatics*, 22(14):e252.

[76] Kim, E.-G. and Klein, M. L. (2004). Unknotting of a polymer strand in a melt. *Macromolecules*, 37(5):1674–1677.

[77] Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., Zhang, M. Q., Lobanenkov, V. V., and Ren, B. (2007). Analysis of the vertebrate insulator protein ctcf-binding sites in the human genome. *Cell*, 128(6):1231 – 1245.

[78] King, N. P., Jacobitz, A. W., Sawaya, M. R., Goldschmidt, L., and Yeates, T. O. (2010). Structure and folding of a designed knotted protein. *Proceedings of the National Academy of Sciences*, 107(48):20732–20737.

[79] King, N. P., Yeates, E. O., and Yeates, T. O. (2007). Identification of rare slipknots in proteins and their implications for stability and folding. *Journal of Molecular Biology*, 373(1):153 – 166.

[80] Klenin, K., Frank-Kamenetskii, M., and Langowski, J. (1995). Modulation of intramolecular interactions in superhelical dna by curved sequences: a monte carlo simulation study. *Biophysical Journal*, 68(1):81 – 88.

[81] Koniaris, K. and Muthukumar, M. (1991). Knottedness in ring polymers. *Phys. Rev. Lett.*, 66:2211–2214.

[82] Kremer, K. and Grest, G. S. (1990). Dynamics of entangled linear polymer melts molecular dynamics simulation. *The Journal of Chemical Physics*, 92(8):5057–5086.

[83] Lim, N. C. and Jackson, S. E. (2015). Mechanistic insights into the folding of knotted proteins in vitro and in vivo. *Journal of Molecular Biology*, 427(2):248 – 258.

[84] Liu, L. F., Liu, C.-C., and Alberts, B. M. (1980). Type ii dna topoisomerases: Enzymes that can unknot a topologically knotted dna molecule via a reversible double-strand break. *Cell*, 19(3):697 – 707.

[85] Lua, R. C. and Grosberg, A. Y. (2006). Statistics of knots, geometry of conformations, and evolution of proteins. *PLOS Computational Biology*, 2(5):1–8.

[86] Macfarlane, R. J., Lee, B., Jones, M. R., Harris, N., Schatz, G. C., and Mirkin, C. A. (2011). Nanoparticle superlattice engineering with dna. *Science*, 334(6053):204–208.

[87] Machon, T. and Alexander, G. P. (2013). Knots and nonorientable surfaces in chiral nematics. *Proceedings of the National Academy of Sciences*, 110(35):14174–14179.

[88] Mallam, A. L. (2009). How does a knotted protein fold? *FEBS Journal*, 276(2):365–375.

[89] Mallam, A. L. and Jackson, S. E. (2005). Folding studies on a knotted protein. *Journal of Molecular Biology*, 346(5):1409 – 1421.

[90] Mallam, A. L. and Jackson, S. E. (2007). A comparison of the folding of two knotted proteins: Ybea and yibk. *Journal of Molecular Biology*, 366(2):650 – 665.

[91] Mallam, A. L., Morris, E. R., and Jackson, S. E. (2008a). Exploring knotting mechanisms in protein folding. *Proceedings of the National Academy of Sciences*, 105(48):18740–18745.

[92] Mallam, A. L., Onuoha, S. C., Grossmann, J. G., and Jackson, S. E. (2008b). Knotted fusion proteins reveal unexpected possibilities in protein folding. *Molecular Cell*, 30(5):642 – 648.

[93] Mansfield, M. L. (1994). Are there knots in proteins? *Nat Struct Mol Biol*, 1(4):213–214.

[94] Mansfield, M. L. (1997). Fit to be tied. *Nat Struct Mol Biol*, 4(3):166–167.

[95] Mansfield, M. L. and Douglas, J. F. (2010). Properties of knotted ring polymers. i. equilibrium dimensions. *The Journal of Chemical Physics*, 133(4):044903.

[96] Mao, C., Sun, W., and Seeman, N. C. (1997). Assembly of borromean rings from dna. *Nature*, 386(6621):137–138.

[97] Marcone, B., Orlandini, E., Stella, A. L., and Zonta, F. (2007). Size of knots in ring polymers. *Phys. Rev. E*, 75:041105.

[98] Marenduzzo, D., Micheletti, C., Orlandini, E., and Sumners, D. W. (2013). Topological friction strongly affects viral dna ejection. *Proceedings of the National Academy of Sciences*, 110(50):20081–20086.

[99] Marenduzzo, D., Orlandini, E., Stasiak, A., Sumners, D. W., Tubiana, L., and Micheletti, C. (2009). Dna–dna interactions in bacteriophage capsids are responsible for the observed dna knotting. *Proceedings of the National Academy of Sciences*, 106(52):22269–22274.

[100] Marko, J. F. and Cocco, S. (2003). The micromechanics of dna. *Physics World*, 16(3):37.

[101] Matthews, K. (1992). Dna looping. *Microbiological reviews*, 56(1):123–136.

[102] Matthews, R., Louis, A. A., and Yeomans, J. M. (2010). Effect of topology on dynamics of knots in polymers under tension. *EPL (Europhysics Letters)*, 89(2):20001.

[103] Meluzzi, D., Smith, D. E., and Arya, G. (2010). Biophysics of knotting. *Annual Review of Biophysics*, 39(1):349–366. PMID: 20192771.

[104] Micheletti, C., Di Stefano, M., and Orland, H. (2015). Absence of knots in known rna structures. *Proceedings of the National Academy of Sciences*, 112(7):2052–2057.

[105] Micheletti, C., Marenduzzo, D., and Orlandini, E. (2011). Polymers with spatial or topological constraints: Theoretical and computational results. *Physics Reports*, 504(1):1 – 73.

[106] Micheletti, C., Marenduzzo, D., Orlandini, E., and Summers, D. W. (2006). Knotting of random ring polymers in confined spaces. *The Journal of Chemical Physics*, 124(6):064903.

[107] Micheletti, C., Marenduzzo, D., Orlandini, E., and Sumners, D. (2008). Simulations of knotting in confined circular {DNA}. *Biophysical Journal*, 95(8):3591 – 3599.

[108] Micheletti, C. and Orlandini, E. (2012). Knotting and metric scaling properties of dna confined in nano-channels: a monte carlo study. *Soft Matter*, 8:10959–10968.

[109] Micheletti, C. and Orlandini, E. (2014). Knotting and unknotting dynamics of dna strands in nanochannels. *ACS Macro Letters*, 3(9):876–880.

[110] Michels, J. P. J. and Wiegel, F. W. (1986). On the topology of a polymer ring. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 403(1825):269–284.

[111] Miguel A. Soler, A. N. and Faisca, P. F. N. (2014). Effects of knot type in the folding of topologically complex lattice proteins. *The Journal of Chemical Physics*, 141(025101):1–10.

[112] Millett, K., Dobay, A., and Stasiak, A. (2005). Linear random knots and their scaling behavior. *Macromolecules*, 38(2):601–606.

[113] Moore, N. T., Lua, R. C., and Grosberg, A. Y. (2004). Topologically driven swelling of a polymer loop. *Proceedings of the National Academy of Sciences of the United States of America*, 101(37):13431–13435.

[114] Najafi, S. and Potestio, R. (2015). Folding of small knotted proteins: Insights from a mean field coarse-grained model. *The Journal of Chemical Physics*, 143(24):243121.

[115] Najafi, S., Tubiana, L., Podgornik, R., and Potestio, R. (2016). Chirality modifies the interaction between knots. *EPL (Europhysics Letters)*, 114(5):50007.

[116] Noel, J. K., Onuchic, J. N., and Sulkowska, J. I. (2013). Knotting a protein in explicit solvent. *The Journal of Physical Chemistry Letters*, 4(21):3570–3573.

[117] Noel, J. K., Sułkowska, J. I., and Onuchic, J. N. (2010). Slipknotting upon native-like loop formation in a trefoil knot protein. *Proceedings of the National Academy of Sciences*, 107(35):15403–15408.

[118] Olavarrieta, L., Martínez-Robles, M. L., Hernández, P., Krimer, D. B., and Schvartzman, J. B. (2002). Knotting dynamics during dna replication. *Molecular Microbiology*, 46(3):699–707.

[119] Orlandini, E., Stella, A. L., and Vanderzande, C. (2010). Geometry and topology of knotted ring polymers in an array of obstacles. *Phys. Rev. E*, 82:050804.

[120] Orlandini, E. and Whittington, S. G. (2007). Statistical topology of closed curves: Some applications in polymer physics. *Rev. Mod. Phys.*, 79:611–642.

[121] Phillips, R., Kondev, J., Theriot, J., and Garcia, H. (2012). *Physical biology of the cell*. Garland Science.

[122] Plimpton, S. (1995). Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 117(1):1 – 19.

[123] Podtelezhnikov, A. A., Cozzarelli, N. R., and Vologodskii, A. V. (1999). Equilibrium distributions of topological states in circular dna: Interplay of supercoiling and knotting. *Proceedings of the National Academy of Sciences*, 96(23):12974–12979.

[124] Poier, P., Likos, C. N., and Matthews, R. (2014). Influence of rigidity and knot complexity on the knotting of confined polymers. *Macromolecules*, 47(10):3394–3400. PMID: 24882882.

[125] Potestio, R., Micheletti, C., and Orland, H. (2010). Knotted vs. unknotted proteins: Evidence of knot-promoting loops. *PLOS Computational Biology*, 6(7):1–10.

[126] Potestio, R. and Tubiana, L. (2016). Discretized knot motion on a tensioned fiber induced by transverse waves. *Soft Matter*, 12:669–673.

[127] Reith, D., Cifra, P., Stasiak, A., and Virnau, P. (2012). Effective stiffening of dna due to nematic ordering causes dna molecules packed in phage capsids to preferentially form torus knots. *Nucleic Acids Research*, 40(11):5129.

[128] Richard Courant, F. J. (1999). *Introduction to Calculus and Analysis I*. Springer Berlin Heidelberg.

[129] Rosa, A., Di Ventra, M., and Micheletti, C. (2012). Topological jamming of spontaneously knotted polyelectrolyte chains driven through a nanopore. *Phys. Rev. Lett.*, 109:118301.

[130] Rosa, A. and Everaers, R. (2008). Structure and dynamics of interphase chromosomes. *PLOS Computational Biology*, 4(8):1–10.

[131] Rosa, A., Orlandini, E., Tubiana, L., and Micheletti, C. (2011). Structure and dynamics of ring polymers: Entanglement effects because of solution density and ring topology. *Macromolecules*, 44(21):8668–8680.

[132] Rothemund, P. W. K. (2006). Folding dna to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302.

[133] Roya Zandi, Y. K. and Kardar, M. (2003). Entropic competition between knots and slip-links. *ARI, Bull. ITU*, 53(6):6.

[134] Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteche, C. M., Filippova, G. N., Baliga, N. S., Aebersold, R., Ranish, J. A., and Krumm, A. (2008). Ctcf physically links cohesin to chromatin. *Proceedings of the National Academy of Sciences*, 105(24):8309–8314.

[135] Rybenkov, V. V., Cozzarelli, N. R., and Vologodskii, A. V. (1993). Probability of dna knotting and the effective diameter of the dna double helix. *Proceedings of the National Academy of Sciences*, 90(11):5307–5311.

[136] Saccà, B. and Niemeyer, C. M. (2012). Dna origami: The art of folding dna. *Angewandte Chemie International Edition*, 51(1):58–66.

[137] Saitta, A. M., Soper, P. D., Wasserman, E., and Klein, M. L. (1999). Influence of a knot on the strength of a polymer strand. *Nature*, 399(6731):46 – 48.

[138] Schoeffler, A. J. and Berger, J. M. (2008). Dna topoisomerases: harnessing and constraining energy to govern chromosome topology. *Quarterly Reviews of Biophysics*, 41(1):41–101.

[139] Seeman, N. C. (1998). Nucleic acid nanostructures and topology. *Angewandte Chemie International Edition*, 37(23):3220–3238.

[140] Seeman, N. C. (2003). Dna in a material world. *Nature*, 421(6921):427–431.

[141] Shimamura, M. K. and Deguchi, T. (2000). Characteristic length of random knotting for cylindrical self-avoiding polygons. *Physics Letters A*, 274(5–6):184 – 191.

[142] Stasiak, A., Katritch, V., Bednar, J., Michoud, D., and Dubochet, J. (1996). Electrophoretic mobility of dna knots. *Nature*, 384(6605):122 – 122.

[143] Sułkowska, J. I., Noel, J. K., Ramírez-Sarmiento, C. A., Rawdon, E. J., Millett, K. C., and Onuchic, J. N. (2013). Knotting pathways in proteins. *Biochemical Society Transactions*, 41(2):523–527.

[144] Sumners, D. W. and Whittington, S. G. (1988). Knots in self-avoiding walks. *Journal of Physics A: Mathematical and General*, 21(7):1689.

[145] Sułkowska, J. I., Sułkowski, P., and Onuchic, J. (2009). Dodging the crisis of folding proteins with knots. *Proceedings of the National Academy of Sciences*, 106(9):3119–3124.

[146] Sułkowska, J. I., Sułkowski, P., Szymczak, P., and Cieplak, M. (2008). Stabilizing effect of knots on proteins. *Proceedings of the National Academy of Sciences*, 105(50):19714–19719.

[147] Taylor, W. R. (2000). A deeply knotted protein structure and how it might fold. *Nature*, 406(6798):916–919.

[148] Taylor, W. R. and Lin, K. (2003). Protein knots: A tangled problem. *Nature*, 421(6918):25–25.

[149] Trefz, B., Siebert, J., and Virnau, P. (2014). How molecular knots can pass through each other. *Proceedings of the National Academy of Sciences*, 111(22):7948–7951.

[150] Tsurusaki, K. and Deguchi, T. (1995). Fractions of particular knots in gaussian random polygons. *Journal of the Physical Society of Japan*, 64(5):1506–1518.

[151] Tubiana, L. (2014). Computational study on the progressive factorization of composite polymer knots into separated prime components. *Phys. Rev. E*, 89:052602.

[152] Tubiana, L., Orlandini, E., and Micheletti, C. (2011). Probing the entanglement and locating knots in ring polymers: A comparative study of different arc closure schemes. *Progress of Theoretical Physics Supplement*, 191:192.

[153] Tubiana, L., Rosa, A., Fragiacomo, F., and Micheletti, C. (2013). Spontaneous knotting and unknotting of flexible linear polymers: Equilibrium and kinetic aspects. *Macromolecules*, 46(9):3669–3678.

[154] Velankar, S., Best, C., Beuth, B., Boutselakis, C. H., Cobley, N., Sousa Da Silva, A. W., Dimitropoulos, D., Golovin, A., Hirshberg, M., John, M., Krissinel, E. B., Newman, R., Oldfield, T., Pajon, A., Penkett, C. J., Pineda-Castillo, J., Sahni, G., Sen, S., Slowley, R., Suarez-Uruena, A., Swaminathan, J., van Ginkel, G., Vranken, W. F., Henrick, K., and Kleywegt, G. J. (2010). Pdbe: Protein data bank in europe. *Nucleic Acids Research*, 38(suppl$_1$) : *D*308.

[155] Virnau, P., Kantor, Y., and Kardar, M. (2005). Knots in globule and coil phases of a model polyethylene. *Journal of the American Chemical Society*, 127(43):15102–15106. PMID: 16248649.

[156] Virnau, P., Mallam, A., and Jackson, S. (2011). Structures and folding pathways of topologically knotted proteins. *Journal of Physics: Condensed Matter*, 23(3):033101.

[157] Virnau, P., Mirny, L. A., and Kardar, M. (2006). Intricate knots in proteins: Function and evolution. *PLOS Computational Biology*, 2(9):1–6.

[158] Virnau, P., Rieger, F. C., and Reith, D. (2013). Influence of chain stiffness on knottedness in single polymers. *Biochemical Society Transactions*, 41(2):528–532.

[159] Vologodskii, A. (2006). Brownian dynamics simulation of knot diffusion along a stretched {DNA} molecule. *Biophysical Journal*, 90(5):1594 – 1597.

[160] Vologodskii AV, Lukashin AV, F.-K. M. A. V. (1994). The knot problem in statistical mechanics of polymer chains. *Sov Phys-JETP*, 39(4):1059–1063.

[161] Wallin, S., Zeldovich, K. B., and Shakhnovich, E. I. (2007). The folding mechanics of a knotted protein. *Journal of Molecular Biology*, 368(3):884 – 893.

[162] Weber, C., Stasiak, A., Rios, P. D. L., and Dietler, G. (2006). Numerical simulation of gel electrophoresis of {DNA} knots in weak and strong electric fields. *Biophysical Journal*, 90(9):3100 – 3105.

[163] Weeks, J. D., Chandler, D., and Andersen, H. C. (1971). Role of repulsive forces in determining the equilibrium structure of simple liquids. *The Journal of Chemical Physics*, 54(12):5237–5247.

[164] Wüst, T., Reith, D., and Virnau, P. (2015). Sequence determines degree of knottedness in a coarse-grained protein model. *Phys. Rev. Lett.*, 114:028102.

[165] Čopar, S., Tkalec, U., Muševič, I., and Žumer, S. (2015). Knot theory realizations in nematic colloids. *Proceedings of the National Academy of Sciences*, 112(6):1675–1680.

[166] Škrbić, T., Micheletti, C., and Faccioli, P. (2012). The role of non-native interactions in the folding of knotted proteins. *PLOS Computational Biology*, 8(6):1–12.