Scuola Internazionale Superiore di Studi Avanzati - Trieste

DOCTORAL THESIS

Static and dynamic properties of knotted biopolymers: from bulk to nanochannels and nanopores

Author:
Antonio SUMA

Supervisor:
Cristian MICHELETTI

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy

Ph.D. course in Physics and chemistry of biological systems
Molecular and Statistical Biophysics group

October 12, 2017

SISSA - Via Bonomea 265 - 34136 TRIESTE - ITALY

DOCTORAL THESIS

Static and dynamic properties of knotted biopolymers: from bulk to nanochannels and nanopores

Author:
Antonio SUMA

Supervisor:
Cristian MICHELETTI

*A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy*

October 12, 2017

"Disce Quasi Semper Victurus, Vive Quasi Cras Moriturus."

Anonymus

Acknowledgements

My first and deepest thanks go to my supervisor Cristian, who is not only a great mentor, but also a wonderful person. I feel privileged to have passed on to me his vast experience and knowledge, which I treasure the most. He is the last of a long list of teachers I encountered in my life: teachers of science, teachers of arts and teachers of life. Giuseppe, Leticia, Enzo, Fulvio, Giancarlo, Valfrido,... , thanks to all of you for making me the person I am today.

I want to thank all the other professors of the SBP group, who contributed substantially to my growth, and made the atmosphere in the SBP group extremely enjoyable.

I want to thank my colleagues and friends of my Phd year: Lorenzo, Richard, Elena, Negar and my officemates too: Matteo, Mattia, Lucia, Guido. We shared countless moments, conversations, and life experiences along the way of these four years.

There have been so many friends I encountered here in SISSA and in Trieste, that I fear I will miss to thank all of them (and most likely I will). Everyone has its special reason to be thanked: trips shared together, fun moments and even simple conversations. Thanks to Edo, Maja, Luca, Alejandro, Marco, Massimo, Emma, Sandro, Angelo, Andrea, Francesca, Jessica, Raffaello, Francesco, Giuseppe, Giovanni, Daniele, Nika, Simon, Sabine, Maria, Matic, Caterina, Antonio, Diego,...

Thanks to the friends I have around Italy and around the world. A special thanks goes to Nicola, who doesn't miss a single occasion to involve me in strange and fun trips, and to Cristian and Max. They are still my best friends after such a long time.

Finally, my family: mom, dad and Rossella. I am grateful for your constant care, patience, help and love; you are and you always be my reference point, my support and my safe haven. Thanks Chiara, you blessed my life with your presence and your love.

I feel really lucky to have all of you around me.

Contents

For my beloved.

Introduction

Ropes or yarns, especially when disorderly packed, are prone to develop knots. Polymers are no exception to this rule and, in fact, rigorous mathematical results have been proved regarding the "statistical necessity" that sufficiently-long circular chains are knotted. The current surge of scientific interest in knotted polymers, and especially biopolymers, is prompted by the need to understand the profound implications that these forms of entanglement can have on the mechanical, dynamical and conformational properties of polymer chains.

For proteins, for instance, a long standing problem is how exactly the knotting properties of naturally-occurring proteins differ from those of general, non-specific, polymer models. This question has, in fact, motivated studies in several directions, from surveying and classifying systematically the repertoire of knots in peptide chains, to establishing the details of the folding route. For genomic DNA instead, it has long been known that it can be highly entangled due to the high packing degree that it attains in all organisms, from viruses to eukaryotic cells. In this case, the advent of single-molecule manipulation techniques, which are routinely applied to DNA filaments of various length, has opened new, and still largely unexplored, perspectives for detecting or controlling the spontaneous knotting properties of DNA.

In this thesis, I will use theoretical and computational techniques to tackle various aspects of the aforementioned issues.

In Chapter 1, I will provide a primer on knots, which sets a reference for concepts and methods used in subsequent chapters. I will in particular present a small resume of knot theory, focusing mainly on notions useful in our context. Afterwards, I will introduce some of the computational techniques used to detect and pinpoint knots along closed and open chains.

In Chapter 2, I will discuss our recent survey of the entire protein data bank, where we searched for all instances of knotted protein chains. The analysis yielded an up-to-date information about the overall knotting probability, the repertoire of knot types, as well as insight on the length and sequence position of knots in peptide chains.

In Chapter 3, I will use a general polyelectrolyte chain model, mapped to DNA, to study the dynamical mechanisms governing knot formation when DNA is confined inside a nanopore channel with size compatible with the DNA persistence length. I will shown that the deep looping and back-folding of the chain ends will be responsible for the knot formation and destruction. Upon increasing the chain length, the knotting probability of DNA increases due to the growing time a knot can diffuse alongside the chain. Instead unknotted lifetimes level off to a constant because they are ruled by the backfolding process.

In chapter 4, I will present the translocation of flexible and knotted polyelectrolyte chains, parametrized after single-stranded DNA, inside a pore too narrow to allow knot passage. This out-of-equilibrium process, which can affect the polymer translocation in complex and counter-intuitive ways, depends deeply on the knot topology. We tackled the resulting translocation compliance in a simple framework

based on how the pulling force, applied only inside the pore, propagates along and past the knot, and how it is related to the structural properties of different knot types.

In chapter 5, I will discuss the translocation of double-stranded DNA chains through wide nanopores. The study is motivated by a recent experimental breakthrough, for which we provide key insight and explanations for the observed phenomenology.

The material presented in this thesis is largely based on the following published papers:

- Jackson, S. E. Suma, A. Micheletti, C. "How to fold intricately: using theory and experiments to unravel the properties of knotted proteins". *Current Opinion in Structural Biology* **42** (2017), pp. 6–14.

- Suma, A. Orlandini, E. Micheletti, C. "Knotting dynamics of DNA chains of different length confined in nanochannels". *Journal of Physics: Condensed Matter* **27.35** (2015), p. 354102.

- Suma, A. Rosa, A. Micheletti, C. "Pore Translocation of Knotted Polymer Chains: How Friction Depends on Knot Complexity". *ACS Macro Letters* **4.12** (2015), pp. 1420–1424.

- Suma, A. Micheletti, C. "Pore translocation of knotted DNA rings". *Proceedings of the National Academy of Sciences* **114.15** (2017), E2991–E2997.

In addition to these published studies I have been involved in two more projects for which we are writing up the results in the following manuscripts that we expect to submit by the end of November 2017:

- Suma, A. Di Stefano, M. Micheletti, C. "Long lived folded states in polyelectrolytes subject to strong electric fields".

- Coronel, L. Suma, A. Micheletti, C. "Dynamics of highly knotted supercoiled DNA".

Chapter 1

Knots in polymer physics

1.1 Introduction

Polymers are filamentous molecules composed by a succession of modular units, or monomers, bound by covalent bonds. The monomers can be strung together in linear, circular or branched polymer structures.

Mayans were one of the first recorded civilizations that exploited polymers to produce various utensils, using the latex extracted from the rubber tree. However, it was only in 1920 that the chemist Hermann Staudinger inferred the existence of materials formed by many covalently-bonded subunits, in order to explain the properties of molecules with high molar masses. Thereafter, scientists developed various chemical reactions to synthetize polymers with different chemical and physical properties, and brought about an important technological revolution in the twentieth century. Today we are all familiar with the ubiquity of synthetic organic polymers, such as polyethylene, polypropylene, polystyrene, that surround us in the guise of plastic objects of various shapes and sizes.

The key molecules of life are also polymeric. They are mainly polysaccharides, with forms ranging from linear to highly branched, proteins and RNA, which are typically linear, and DNA, that can be organized in both linear and circular forms [1, 2].

Polymers, similarly to sufficiently long ropes and yarn that are carelessly stowed away, can be prone to develop various forms of entanglement, including knots.

As a matter of facts, in a celebrated mathematical conjecture, Frisch, Wasserman and Delbruck [3, 4] posited in the early 1960's that asymptotically-long random walks had to be knotted except for a set of measure zero. In 1988, nearly 30 years after its formulation, Sumners and Whittington proved this conjecture [5, 6] by demonstrating that long self-avoiding walks on a cubic lattice are bound to feature tight knots. The self-avoiding constraint prevents these tight knots to be threaded through and hence be undone by other parts of the chains. This demonstrated the "statistical necessity" of knots occurrence in sufficiently long polymers first formalized by Frisch and colleagues.

Several pioneering experiments and structural surveys demonstrated that knots can appear in biopolymers such as RNA [7], proteins [8–10] and DNA [11–15]. These discoveries brought forth several questions concerning how knots can be selected by evolution [9, 16, 17] and how they reverberate *in-vivo* [18].

In the next sections I will provide a self-contained primer on knots that will serve as introduction to the results of the following chapters, which are focussed mainly on knots in proteins and DNA. I will start with the definition and characterization of knots in closed, ring-like chains, and then discuss how the concepts extend to the case of open, linear chains. The algorithms that can be used to detect and pinpoint knots in both cases will be briefly described.
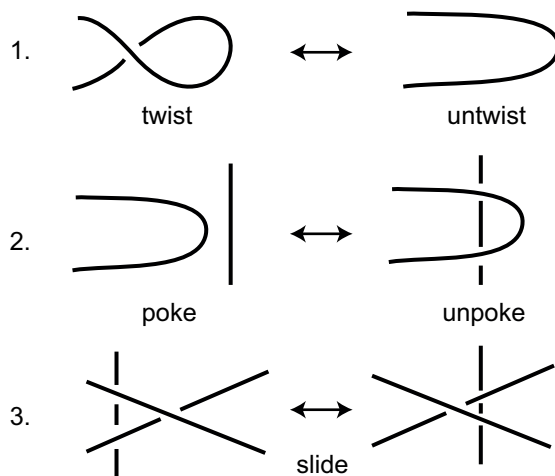
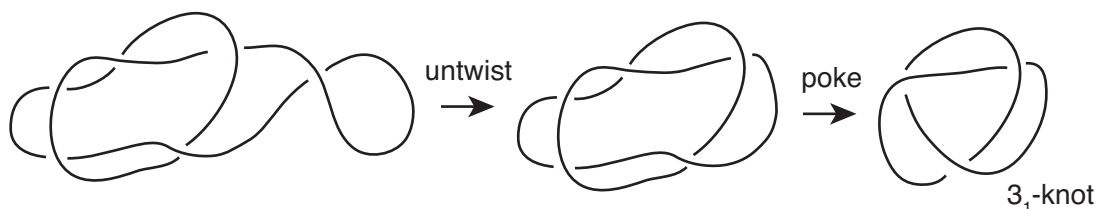FIGURE 1.1: Reidemeister moves, conserving the knot type on a knot diagram.

FIGURE 1.2: Simplification of a knot diagram into its minimal representation, through a combination of the available Reidemeister moves. Other moves cannot further reduce the number of crossings.

1.2 Knot theory

In topology, knots are embeddings of closed curves in Euclidean 3-dimensional space $\mathbb{R}^3$, whose properties are preserved under ambient isotopies. These are continuous, non singular deformations of the curve that do not lead to self-crossings. The deformations include bending, twisting, stretching but not cutting and gluing. Curves of different geometries are topologically equivalent if they can be transformed one into the other through ambient isotopies, and these classes of equivalence define the different knot types.

A convenient way to represent and visualize a knot is to project the curve on a plane, tracking the over- and under-passes. In such representations, ambient isotopies can be obtained by combining a suitable succession of the three Reidemeister moves [19], a basis set of deformations shown in Fig. 1.1. Two closed curves are therefore equivalent if it is possible to pass from one planar projection to the other using a combination of Reidemeister moves. Furthermore, if any set of Reidemeister moves cannot further reduce the number of crossings of a given knot, these crossings will be termed essentials, and the diagram will be in its minimal representation, see Fig. 1.2 for an example.

It is always possibile to introduce two or more knots in the same ring, producing a composite knot or connected sum of knots, see Fig. 1.3 for an example. Instead, knots that cannot be further decomposed in terms of a non-trivial connected sum are called prime knots.
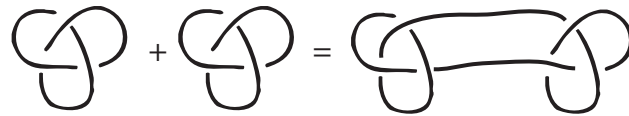
FIGURE 1.3: Example of composite knot construction.

Prime knot are traditionally labelled with n_i, where n is the number of essential crossings, and i is a conventional index used to distinguish different knot types with the same number of essential crossings n [20]. For example, the trefoil knot appearing in Fig. 1.2 is indicated with 3_1. Composite knots formed by two or more knots $n_i, m_j, ..., l_k$ are instead labelled with $n_i \# m_j \# ... \# l_k$.
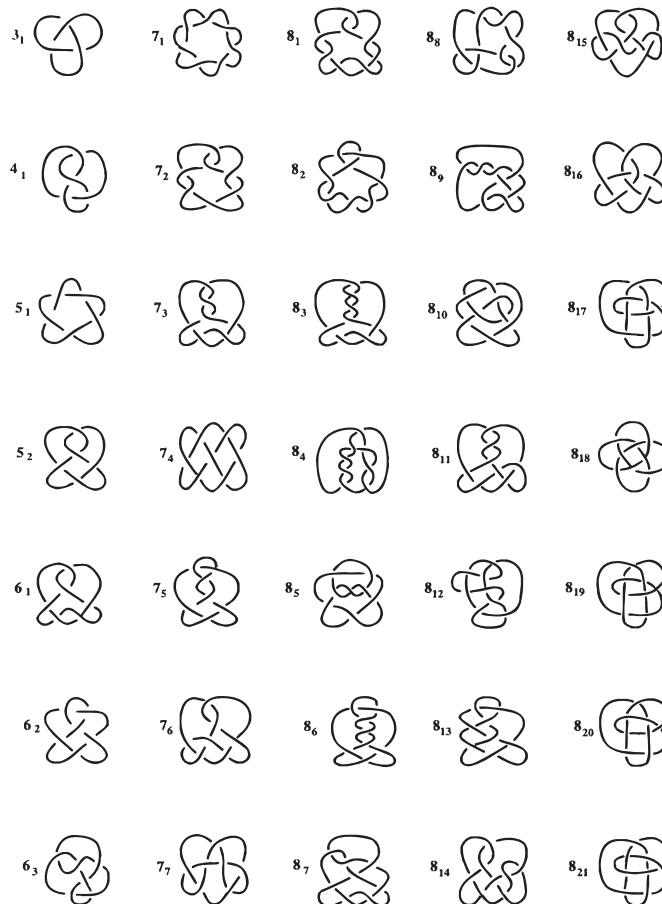
FIGURE 1.4: Knot table up to eight crossing, from [21].

Fig. 1.4 shows all prime knots up to $n = 8$ crossings, in their minimal representations, excluding the trivial case with zero crossings 0_1, which corresponds to a circle.

The simplest non-trivial knot is the trefoil knot or 3_1-knot, which has three essential crossings. The 3_1 is a chiral knot, because it has in its minimal representation two mirror forms: the left-handed, shown in Fig. 1.4, and the right-handed, and they are not deformable continuously into each other. The second case is a 4_1-knot with four-crossings, also known as figure-of-eight knot. This knot is achiral because it is deformable continuously in its mirrored image. The next two prime knots have five crossing and are both chiral: the pentafoil knot or 5_1 and the 5_2-knot.
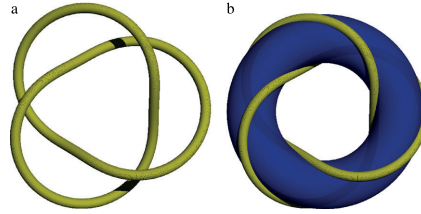
FIGURE 1.5: Torus knot (trefoil 3_1) constructed on the surface of a
torus.
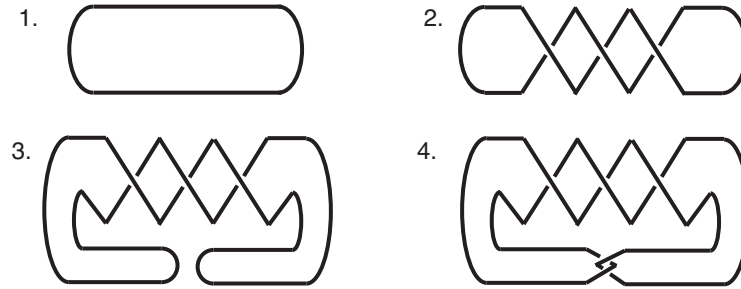
FIGURE 1.6: Twist knot construction steps.

The number of distinct knot types steadily increases with the number of minimal crossings n: we find three knot types with $n = 6$, seven with $n = 7$ and twenty-one with $n = 8$. In fact, the number of knot types for a fixed number of crossings grows approximatively exponentially [22, 23], and they have presently been tabulated up to $n = 16$ crossings [24], for a total amount of about 1.7 million.

Prime knots can also be grouped in families according to various criteria. Two salient families are the torus and twist ones, which will be discussed in subsequent chapters. Torus knots can be obtained by winding a curve a number of times around the equator and the meridians of a torus, see Fig. 1.5, and include $3_1, 5_1, 7_1, 8_{19}, 9_1$. Twist knots are obtained by twisting a closed 0_1 ring repeatedly and then by making a clasp passage of the two end loops, see Fig. 1.6. They include $3_1, 4_1, 5_2, 6_1, 7_2, 8_1$. Note that the simplest knot type 3_1 belongs to both families.

Assigning the knotted state of a curve from one of its possible diagrammatic representations can be a daunting task, even if the knot is a tabulated one. For this purpose, topological invariants —algebraic quantities unaffected by the different geometries in which a given knotted ring presents itself— are commonly used.

In this thesis we will mostly use the Alexander invariant [25], because it is computationally efficient. The algorithm that computes such invariant requires several steps: an orientation is attached to the curve, and then the crossings are classified as positive or negative by a right-hand rule. Afterwards, a matrix is built assigning to each element a monomial, related to the crossing type, and the determinant of the matrix is computed. The result is the so called Alexander polynomial associated to the given topology [26, 27].

The resulting polynomial is independent of the chosen projection and the different geometries the knot can present itself, allowing to distinguish between different knot types, but not different chiral forms. There are however several counterexamples showing that genuinely complex knot types can have the same polynomial of simpler knot types, and even of the unknot. Nevertheless, the cases we shall encounter will be simply enough to not display such ambiguity. There also exist other

polynomials, such as HOMFLY, that can discriminate these cases, although computationally more demanding.

1.3   Defining and detecting knots in open chains

Suppose to tie several knots on a linear rope and wish to quantify its degree of entanglement. This is a task that eludes a strict mathematical formulation, because no matter the complexity of the tangle, there will exists always a set of continuous deformations, not involving strand crossings, that transform it into a linear, hence unknotted, segment.

Nevertheless, we are all familiar with the mechanical effects that knots tied on an open chain can have: for example, a knotted fishing line has a much lower rupture force than an unknotted one [28]. Hence, these kind of knots, even thought they lack a strict mathematical definition, are relevant and are usually referred to as physical knots.

In the microscopic domain, as well as the macroscopic one, physical knots can similarly affect the chain's metric and dynamical properties, and can be long-lived in sufficiently long liner polymers, as we will see in Chapter 3.

An intuitive way to characterize the degree of entanglement of an open chain is by joining the two ends with an auxiliary arc. This operation fixes the chain's topological state and traps the physical entanglement on a ring, allowing to measure the Alexander invariant.

In principle, the way the two ends are joined together can give different knot types. The procedures listed in the following are designed to be simply and transparent:

- Historically, the first procedure conceived was the stochastic closure to infinity [9]: the two ends are joined together, with two segments, on a random point taken on a sphere, with the sphere centered around the chain's center of mass and with radius much larger than its gyration radius, see Fig. 1.7a. Different random points will generally yield different types of knots, so for this reason one can take the most common topology out of the many closures as the representative one.

- The second procedure, called minimally-interfering closure [29], compares the arc lengths introduced by two different closure schemes. The first scheme joins the two ends with a straight line, and is hence called direct bridging, see Fig. 1.7b. The second scheme prolongs the ends to their respective closest point at the convex hull and then joins them together with an arc at infinity, see Fig. 1.7c. The scheme with the shortest total arc length (excluding the closure at infinity for the second scheme) is selected, in order to introduce the least amount of spurious crossings.

Both methods have advantages and disadvantages: the stochastic closure provides a more complete statistics on the possible topologies of the chain, but needs the calculation of many knot invariants, while the minimally-interfering closure requires a single calculation and generally gives the same dominant type [29]. We will use the latter method for computational effectiveness.

Closure to infinity | Minimally-interfering closure

a)

b) Direct bridging

c) Closure at the convex hull
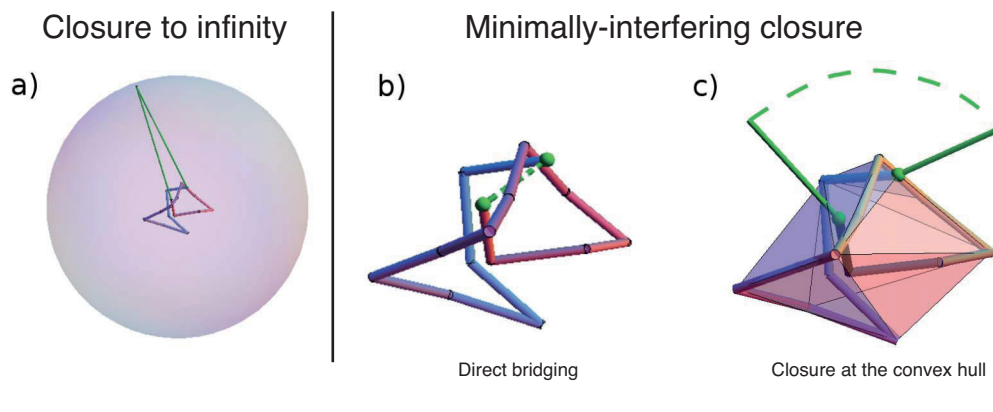
FIGURE 1.7: Examples of different closure schemes, from [29]. (a) Stochastic closure at infinity, (b)-(c) Minimally-interfering closure, choosing the shortest arc length between the direct bridging (b) and the closure at the convex hull (c).

1.4   Knot localization schemes

The closure schemes introduced in the previous section can be used to pinpoint where exactly the knot resides along the chain. The idea is to find the minimal open portion of the chain with the same topology as the whole chain, and can be applied both to open and closed chains.

Two search schemes are possibile [29–32]:

- The bottom-up scheme starts by considering all the possibile arcs with a minimal length (the shortest that can contain a knot after closure) along the chain. Then for each arc, after closure, the topological state is checked to be the same as the whole chain, and that its complementary arc is instead unknotted. If no arc satisfies this criteria, the arc length is increased. The procedure ends when one arc satisfies the aforementioned criteria, and it will be considered the shortest knotted arc of the whole chain.

- The top-down approach instead considers all the possibile arcs of given length l, starting from the maximum possible one (the whole chain), and identifies

a)   b)   c)
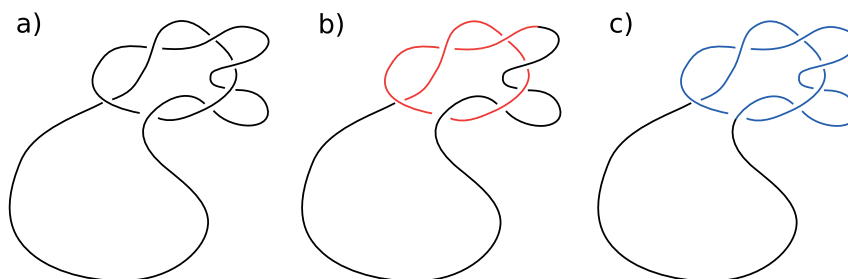
FIGURE 1.8: Knot localization schemes, from [29], applied over a trefoil knotted configuration a). b) The bottom-up approach detects the shortest length over which topological complexity appears. c) The top-down approach detects the shortest length to which topology can be maintained. The two approaches identify in this case a different knotted portion.

the arcs that have the same topology as the entire chain. Only the arcs satisfying this criterion will be then subdivided in smaller arcs of length $l - 1$, while the others will be discarded. The surviving arcs will be again checked with the same criterium, and the arcs satisfying it will be selected and further subdivided. The procedure ends if, after a subdivision, no one of the arcs has the same topology of the entire chain. Then, the arc that survived the previous iteration will be considered the shortest continuously-knotted portion of the ring.

The knotted arc found with the two procedures might generally differ when the chain is in a compact state, see Fig. 1.8. Since we will look for the shortest possible knotted portion, we will use in our case the bottom-up approach.

Chapter 2

Knotted proteins: an up-to-date survey

2.1 Introduction

Proteins are linear biopolymers that perform essential functions inside living organisms: they can act as cellular scaffolds, can catalyze thousands of metabolic reactions, serve as molecular machines, and in general are key to cell signalling processes.

The amino-acids (a.a.) are the protein monomeric units, assembling in sequence through covalent bonds. There are 20 main types of naturally occurring amino acids, which can be arranged in different sequence orders, thus providing a huge combinatorial space for protein sequences.

The different types of amino acids share a common chemical backbone, and are distinguished by the side-chain or residue group (R-group), see Fig. 2.1, which is attached to a pivoting carbon atom of the backbone (α-carbon or C_α).

Naturally occurring proteins have sequences spanning only a small subset of the 20 amino acids full combinatorial space. Evolutionary pressure has arguably optimized sequences to have a unique native fold that minimizes the protein's free energy. In addition, sequences have been optimized for specific functions and catalytic activities, to fold fast enough to avoid proteosomal degradation, etc.

For decades, the scientific community held a strong belief that proteins had to be knot free on the ground that, if a protein was prone to form knots during the folding process, then it would need to backtrack from these states in order to fold into the correct native state. Knots would therefore be kinetic traps that hinder the folding
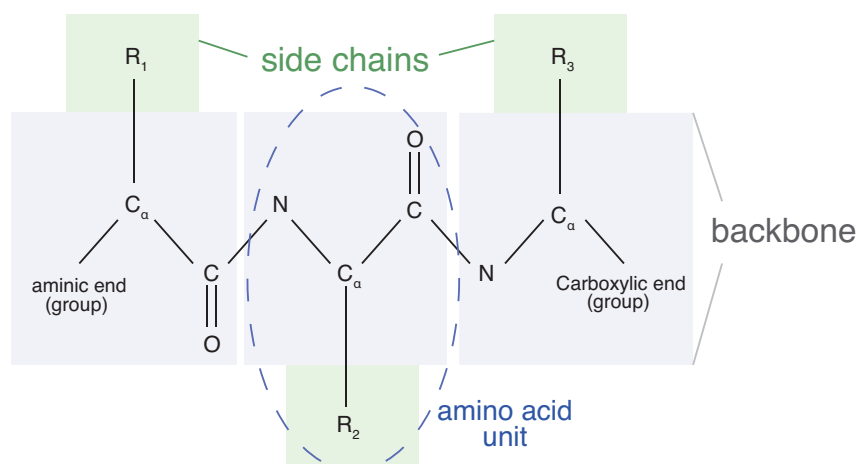
FIGURE 2.1: Schematic composition of a protein.
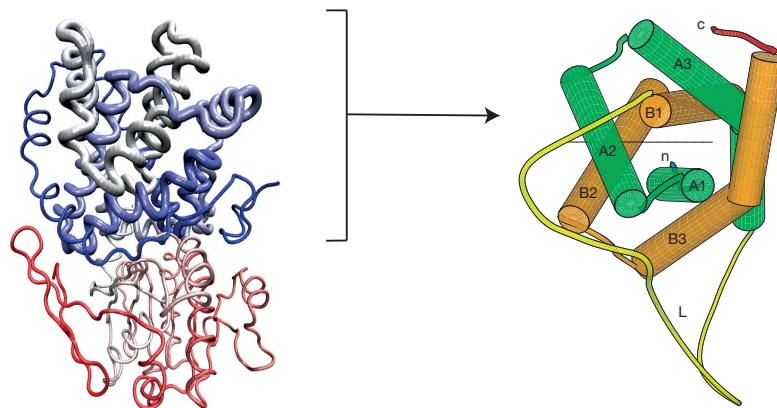
FIGURE 2.2: On the left, the first deeply knotted protein discovered [10]: a figure of eight (4_1) knot with the knot ends distant 245 a.a. from the N-terminus, and 70 a.a. from the C-terminus. On the right, a view of the knotted core, starting from the C-terminus, from [10].

process, which needs instead to be as efficient as possible, in order for the protein to avoid degradation.

However, as more protein structures were released in the PDB, several proteins were eventually found to be knotted. The first documented instance was a carbonic anhydrase [8, 9], found in 1977, featuring a shallow 3_1-knot, with just few residues threading through a wide loop. The second case was an acetohydroxy acid iso-meroreductase, found in 2000, featuring a 4_1 knot [10], see Fig. 2.2. In this case, the knot ends were distant 245 a.a. and 70 a.a. from the termini, providing a compelling evidence that proteins can fold into deeply knotted structures. This discovery posed new questions about the steps that coordinate the folding process [33–36] and about knotted proteins evolutionary and functional putative advantages over non-knotted homologs [17, 37].

The ongoing growth of structures available in the PDB poses the need for a continuous search for new knotted protein families and new knot types. For this reason, we have undertook in 2016 a survey of the PDB in search for all knotted protein structures, their representatives and their similarities or differences with the case of generic polymers.

The main findings of this chapter are based on the recent review article of Ref. [38].

2.2   Knotted proteins representative list construction

Our survey of the PDB involved the following steps, which are also sketched in the flow-chart of Fig. 2.3.

A: Fetch of all protein crystal structures available
First, we downloaded the entire set of PDB structures available as of June 2016, retrieving all the entries containing proteins, which yielded about $5 \cdot 10^5$ individual

Knotted proteins representative list construction
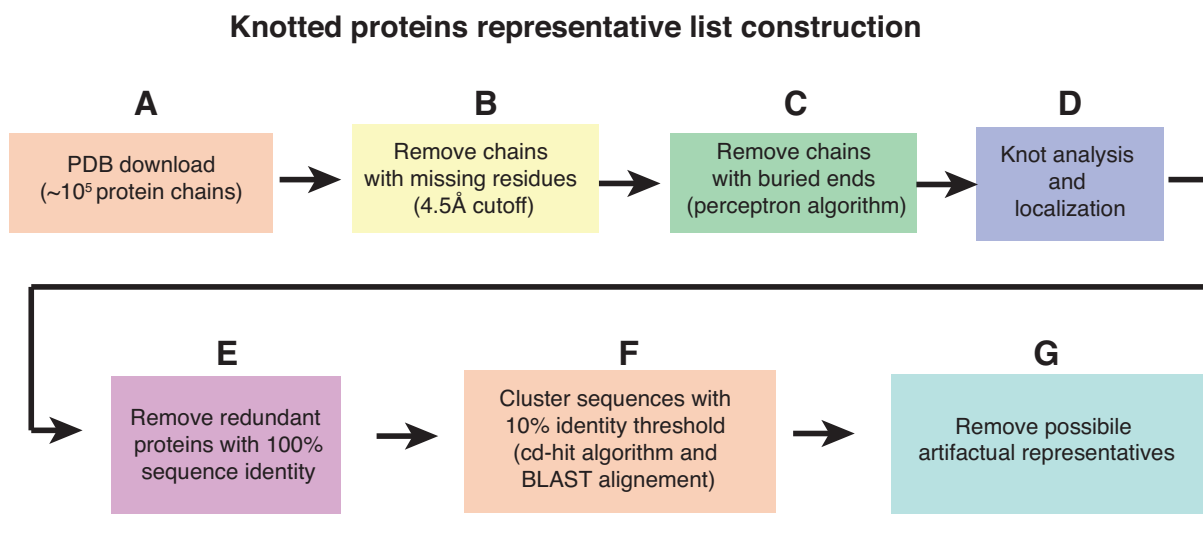
FIGURE 2.3: Block diagram with the knotted protein survey procedure.

peptide chains counting separately all models of structures solved by NMR or X-ray solution scattering.

B: Remove chains with missing residues

For each protein chain, we retained only the α-carbon atoms, as representatives of the protein backbone.

We considered only chains without missing intermediate residues. This check was done by computing the distance between the consecutive α-carbons along the chain. In particular, a residue was considered missing if two consecutive α-carbons were distant more than 4.5Å. This value was selected considering that in the typical *trans* case the average distance is 3.8Å and its standard deviation 0.04Å [39] (the rarer *cis* case has a lower average distance).

We have used the same cutoff distance to check if some structure, such as the cyclotide, had a cyclized backbone, where the N-terminus is covalently bonded to the C-terminus.

C: Remove chains with deeply buried ends

We processed separately chains with a cyclized, circular, backbone, and open ones. In the former case, we can in fact probe the topological state directly by calculating the Alexander determinant. In the latter, instead, we need first to close the chain by connecting the ends with an auxiliary arc, in order to turn it into a ring and then establish its topology by computing the Alexander determinant.

The auxiliary arc, as we discussed in Chapter one, is not uniquely defined, and different topological states might be inferred if we change the arc's position. This is particularly true for proteins, because they are typically compact and might have one or both ends deeply buried inside.

To ensure that the auxiliary arc does not introduce spurious entanglement, we selected only structures with both termini exposed. In this way, the ends can be bridged univocally with an arc at infinity.

Assessing whether a terminal is exposed amounts to finding a plane, passing through the terminal, such that all the amino-acids are included in one of the two semi-space identified by the plane, see Fig. 2.4. This is equivalent to find a vector
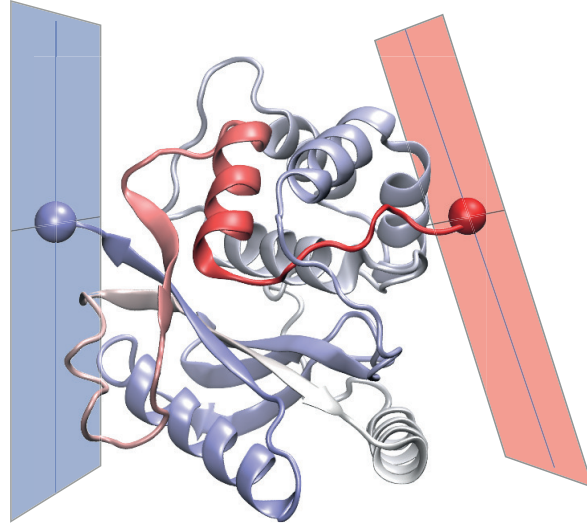
FIGURE 2.4: Protein with exposed ends: for each terminal, there exists
a plane, passing through the terminal, such that all the amino-acids
are included in one of the two semi-space identified by the plane.

$\mathbf{w}$ (the normal of the aforementioned plane) pointing in the opposite direction of all
the distance vectors between the terminal and each α-carbons of the chain.

The existence (or non-existence) of $\mathbf{w}$ can be proved using a linear binary classi-
fier algorithm (perceptron), described in the following.

Suppose we have a protein chain with N residues and where $\mathbf{x}_i$ are the α-carbon
coordinates, with index $i = 0, \ldots N - 1$. Without loss of generality, we now focus on
the first end with index $i = 0$. The vector $\mathbf{w}$ can then be found through the following
procedure:

1. Initialize $\mathbf{w}$ as a random unit vector

2. Compute the projections d_i of all the distances $\mathbf{x_i} - \mathbf{x_0}$ onto $\mathbf{w}$, $d_i = \mathbf{w} \cdot (\mathbf{x_i} - \mathbf{x_0})$
   for $i = 1, \ldots N - 1$

3. Find the largest projection, defined as $d_k = \max\{d_i\}_{i=1,\ldots N-1}$

4. Take as new vector $\mathbf{w} = \mathbf{w} - \epsilon(\mathbf{x_k} - \mathbf{x_0})$, with $\epsilon = 10^{-3}$ and restart from step 2.

The end will be considered exposed if, after a certain maximum number of itera-
tion steps (in our case $10^5 N$), the largest projection yields $d_k > 0$, which means that
all distance vectors, $\mathbf{x_i} - \mathbf{x_0}$, point in the opposite direction of $\mathbf{w}$. Instead, in the case
in which $d_k \leq 0$, the end will be considered buried.

If a separating plane exists, the perceptron algorithm is guaranteed to converge
to the optimal solution through a finite number of iterations proportional to the
largest $(\mathbf{x_i} - \mathbf{x_0})^2$.

The whole procedure can then be applied to terminal $i = N - 1$ to check if it is
exposed or not.

D: Knot analysis and localization

After having identified the protein entries with exposed termini, we proceeded to
bridge their termini using the minimally-interfering closing procedure.

2K0A:A(2)	2K0A:A(3)	2K0A:A(4)	2K0A:A(5)	2K0A:A(7)	2K0A:A(9)	2K0A:A(10)	2K0A:A(12)	2K0A:A(15)
2K0A:A(16)	2K0A:A(19)	2K0A:A(20)	2LEN:A(1)	2LEN:A(2)	2LEN:A(3)	2LEN:A(4)	2LEN:A(6)	2LEN:A(7)
2LEN:A(8)	2LEN:A(9)	2LEN:A(10)	2LEN:A(11)	2LEN:A(12)	2LEN:A(14)	2LEN:A(15)	2LEN:A(17)	2LEN:A(18)
2LEN:A(19)	2LEN:A(20)	3KKX:A(1)	4G0C:A(1)	4Q49:A(1)	4Y0J:A(1)	1A42:A	1AM6:A	1AVN:A
1AZM:A	1BCD:A	1BIC:A	1BN1:A	1BN4:A	1BNM:A	1BNN:A	1BNQ:A	1BNT:A
1BNU:A	1BNV:A	1BNW:A	1BV3:A	1BZM:A	1CA2:A	1CAH:A	1CAI:A	1CAJ:A
1CAK:A	1CAL:A	1CAM:A	1CAY:A	1CAZ:A	1CIL:A	1CIM:A	1CIN:A	1CNG:A
1CNH:A	1CNI:A	1CNJ:A	1CNW:A	1CNX:A	1CNY:A	1CRA:A	1CZM:A	1DMX:A
1DMX:B	1DMY:A	1DMY:B	1EOU:A	1F2W:A	1FLJ:A	1FQN:A	1FQR:A	1FR4:A
1FSN:A	1FSN:B	1FSQ:A	1FSQ:B	1FSR:A	1FSR:B	1FUG:B	1G0E:A	1G0F:A
1G1D:A	1G3Z:A	1G4J:A	1G4O:A	1G45:A	1G46:A	1G48:A	1G52:A	1G53:A
1G54:A	1GZ0:A	1GZ0:B	1GZ0:C	1GZ0:D	1GZ0:E	1GZ0:F	1GZ0:G	1GZ0:H
1H4N:A	1H9N:A	1HCB:A	1HEA:A	1HEC:A	1HUH:A	1I8Z:A	1I9L:A	1I9M:A
1I9N:A	1I9O:A	1I9P:A	1I9Q:A	1I90:A	1I91:A	1IF4:A	1IF5:A	1IF6:A
1IF7:A	1IF8:A	1IF9:A	1J9W:A	1JCZ:B	1JV0:A	1KEQ:A	1KEQ:B	1KWQ:A
1KWR:A	1LG5:A	1LG6:A	1LGD:A	1MUA:A	1NXZ:A	1NXZ:B	1O6D:A	1OKM:A
1OQ5:A	1P7L:A	1P7L:B	1P7L:C	1P7L:D	1QMG:A	1QMG:D	1RAY:A	1RAZ:A
1RG9:A	1RG9:B	1RG9:C	1RG9:D	1RJ5:A	1RJ6:A	1RZA:A	1RZB:A	1RZC:A
1RZD:A	1RZE:A	1T9N:A	1TB0:A	1TBT:A	1TE3:A	1TEQ:A	1TEU:A	1TG3:A
1TG9:A	1TH9:A	1THK:A	1TTM:A	1UGA:A	1UGB:A	1UGC:A	1UGD:A	1UGE:A
1UGF:A	1UGG:A	1URT:A	1V9E:A	1V9E:B	1V9I:A	1VH0:A	1VH0:B	1VH0:C
1VH0:E	1X7O:A	1X7P:A	1XD3:A	1XD3:C	1XEG:A	1XEV:A	1XEV:B	1XEV:C
1XEV:D	1XPZ:A	1XQ0:A	1YDD:A	1YO0:A	1YO1:A	1YO2:A	1YVE:A	1YVE:D
1Z9Y:A	1Z93:A	1Z97:A	1ZE8:A	1ZFQ:A	1ZGE:A	1ZGF:A	1ZH9:A	1ZJR:A
1ZSA:A	1ZSB:A	1ZSC:A	2ABE:A	2AW1:A	2AX2:A	2CA2:A	2CBA:A	2CBB:A
2CBC:A	2CBD:A	2EFV:A	2EGV:A	2EGV:B	2ETL:A	2ETL:B	2EU2:A	2EU3:A
2EZ7:A	2F14:A	2FG6:A	2FG6:B	2FG6:C	2FG6:D	2FG6:E	2FG6:F	2FG7:A
2FG7:B	2FG7:C	2FG7:D	2FG7:E	2FG7:F	2FMG:A	2FMZ:A	2FNK:A	2FNM:A
2FNN:A	2FOQ:A	2FOS:A	2FOV:A	2FOY:B	2FW4:A	2G7M:A	2G7M:B	2G7M:C
2G7M:D	2G7M:E	2G7M:F	2GEH:A	2H4N:A	2H15:A	2HA8:A	2HD6:A	2HKK:A
2HL4:A	2HNC:A	2HOC:A	2ILI:A	2NMX:A	2NMX:B	2NN1:A	2NN1:B	2NN7:A
2NN7:B	2NNG:A	2NNO:A	2NNS:A	2NNV:A	2NWO:A	2NWP:A	2NWY:A	2NWZ:A
2NXR:A	2NXS:A	2NXT:A	2O4Z:A	2O9C:A	2OBV:A	2OSF:A	2OSM:B	2P02:A
2POU:A	2POV:A	2POW:A	2Q1B:A	2Q1Q:A	2Q38:A	2QMM:A	2QMM:B	2QO8:A
2QOA:A	2QP6:A	2RH3:A	2VVA:A	2VVB:A	2WD3:A	2WEG:A	2WEH:A	2WEJ:A
2WEO:A	2X7T:A	2X7U:A	3B4F:A	3BBD:A	3BBD:B	3BBE:A	3BBE:B	3BBH:A
3BBH:B	3BJX:B	3BJX:C	3BL0:A	3BL1:A	3C7P:A	3CAJ:A	3CYU:A	3CZV:A
3CZV:B	3D0N:A	3D0N:B	3D8W:A	3D9Z:A	3D92:A	3D93:A	3DA2:A	3DAZ:A
3DBU:A	3DC3:A	3DC9:A	3DCC:A	3DCS:A	3DCW:A	3DD0:A	3DD8:A	3DV7:A
3DVB:A	3DVC:A	3DVD:A	3EFI:A	3EFT:A	3F4X:A	3F8E:A	3FFP:A	3GZ0:A
3HFP:A	3HKN:A	3HKQ:A	3HKT:A	3HKU:A	3HLJ:A	3HS4:A	3IAI:A	3IAI:B
3IAI:C	3IAI:D	3IBI:A	3IBL:A	3IBN:A	3IBU:A	3IEO:A	3IFW:A	3IGP:A
3IRT:A	3IRT:B	3JXF:A	3JXF:B	3K2F:A	3K7K:A	3K34:A	3KIG:A	3KNE:A
3KOK:A	3KON:A	3KS3:A	3KTY:C	3KVF:A	3KW5:A	3KWA:A	3KZK:A
3L14:A	3M1J:A	3M1K:A	3M1Q:A	3M2N:A	3M2X:A	3M2Y:A	3M2Z:A	3M3X:A
3M5E:A	3M5S:A	3M5T:A	3M14:A	3M40:A	3M67:A	3M96:A	3M98:A	3MDZ:A
3MHC:A	3MHI:A	3MHL:A	3MHM:A	3MHO:A	3ML2:A	3MN6:A	3MNA:A	3MNH:A
3MNI:A	3MNJ:A	3MNK:A	3MNU:A	3MWO:A	3MWO:B	3MYQ:A	3MZC:A	3N0N:A
3N2P:A	3N3J:A	3N4B:A	3N4J:A	3N4K:A	3N4K:B	3NB5:A	3NI5:A	3NJ9:A
3NK7:B	3O7B:A	3OIK:A	3OIL:A	3OIM:A	3OKU:A	3OKV:A	3ONP:A	3OY0:A
3OYQ:A	3OYS:A	3P3H:A	3P3J:A	3P4V:A	3P5A:A	3P5L:A	3P44:A	3P55:A
3P58:A	3PJJ:A	3PO6:A	3PYK:A	3QYK:A	3R16:A	3R17:A	3RG4:A	3RJ7:A
3RLD:A	3RYJ:A	3RYV:A	3RYX:A	3RYY:A	3RYZ:A	3RZ0:A	3RZ1:A	3RZ5:A
3RZ7:A	3RZ8:A	3S8X:A	3S9T:A	3S71:A	3S72:A	3S73:A	3S74:A	3S75:A
3S76:A	3S77:A	3S78:A	3SAP:A	3SAX:A	3SBH:A	3SBI:A	3T5U:A	3T5Z:A
3T82:A	3T83:A	3T84:A	3T85:A	3TMJ:A	3TVN:A	3TVO:A	3U3A:A	3U7C:A
3U45:A	3U47:A	3ULK:A	3ULK:B	3UYN:A	3UYQ:A	3V2J:A	3V2M:A	3V3F:A
3V3G:A	3V3H:A	3V3I:A	3V3J:A	3V5G:A	3V7X:A	3VBD:A	3WJ8:A	3WJ8:B
3WJ8:C	3WJ8:D	3WJ8:E	3WJ8:F	3ZP9:A	3ZQ5:A	4BCW:A	4BF1:A	4BF6:A
4CAC:A	4CND:A	4CNE:A	4CNR:B	4CNR:D	4CNV:A	4CNW:A	4CNW:B	4DM9:A
4DM9:B	4DZ7:A	4DZ9:A	4E3D:A	4E3F:A	4E3G:A	4E3H:A	4E04:A	4E04:B
4E4A:A	4E5Q:A	4E8B:A	4E49:A	4FAK:A	4FIK:A	4FL7:A	4FPT:A	4FRC:A
4FU5:A	4FVN:A	4FVO:A	4G7A:A	4G7A:B	4GL1:A	4GW9:A	4H3Z:B	4HBA:A
4HEW:A	4HEY:A	4HEZ:A	4HF3:A	4HT0:A	4HU1:A	4IDR:A	4IG6:A
4ILX:A	4ITO:A	4ITP:A	4IWZ:A	4JAK:A	4JAK:B	4JAL:A	4JAL:B	4JKJ:A
4JKJ:B	4JS6:A	4JSA:A	4JSS:A	4JSW:A	4JSZ:A	4JWF:A	4JWF:B	4JWG:A
4JWJ:A	4JWJ:B	4K0S:A	4K0T:A	4K1Q:A	4K13:A	4KAP:A	4KGN:C	4KGN:F
4KNI:A	4KNJ:A	4KNM:A	4KPP:A	4KTT:B	4KTT:D	4KTV:B	4KTV:D	4KUW:A
4KUY:A	4KV0:A	4L5U:A	4L5V:A	4L5W:A	4LHI:A	4LP6:B	4LRV:A	4LRV:B
4LRV:C	4LRV:D	4LRV:F	4LRV:G	4LRV:H	4LRV:J	4LRV:O	4LU3:A	4M2R:A
4M2U:A	4M2V:A	4M2W:A	4MDG:A	4MDL:A	4MDM:A	4MLT:A	4MLX:A	4MO8:A
4MTY:A	4N0X:A	4N2X:A	4N2X:B	4N2X:C	4N2X:D	4N2X:E	4N2X:F	4N16:A
4NDN:B	4NDN:D	4ODJ:A	4PXX:A	4PYX:A	4PYY:A	4PZH:A	4Q06:A	4Q6D:A
4Q6E:A	4Q07:A	4Q7P:A	4Q7S:A	4Q7V:A	4Q7W:A	4Q08:A	4Q8X:A	4Q8Y:A
4Q8Z:A	4Q09:A	4Q9Y:A	4Q78:A	4Q81:A	4Q83:A	4Q87:A	4Q90:A	4Q99:A
4QEF:A	4QIY:A	4QIY:B	4QIY:C	4QIY:D	4QJM:A	4QK1:A	4QK2:A	4QK3:A
4QSA:A	4QSB:A	4QSI:A	4QSJ:B	4QTL:A	4QY3:A	4R5B:A	4R59:A	4RFC:A
4RFD:A	4RH2:A	4RIU:A	4RIV:A	4RN4:A	4RUX:A	4RUY:A	4RUZ:A	4WL4:A
4WR7:A	4WR7:B	4WUP:B	4WUQ:A	4WUQ:B	4WW6:A	4X3L:A	4X3L:B	4X3M:A
4X3M:B	4XE1:A	4XFW:A	4XFW:B	4XZ5:A	4XZ5:B	4XZ5:C	4XZ5:D	4Y3I:A
4Y5F:A	4YGF:C	4YGF:E	4YGF:G	4YGJ:A	4YGK:A	4YGL:A	4YGN:A
4YHA:C	4YVI:B	4YVJ:B	4YVK:B	4YVY:A	4YX4:A	4YXI:A	4YXO:A	4YXU:A
4YYT:A	4Z0Q:A	4Z1E:A	4Z1J:A	4Z1K:A	4Z1N:A	4ZAO:A	4ZWX:A	4ZWY:A
4ZWZ:A	4ZX0:A	4ZX1:A	5A1G:A	5A1I:A	5A25:A	5A25:B	5AMD:A	5AMG:A
5AML:A	5BNL:A	5BRV:A	5BRW:A	5BU6:A	5BU6:B	5BYI:A	5CAC:A	5CJF:A
5CLU:A	5DSI:A	5DSJ:A	5DSK:A	5DSL:A	5DSM:A	5DSN:A	5DSO:A	5DSP:A
5DSQ:A	5DSR:A	5E2K:A	5E2S:A	5E28:A	5EH8:A	5EHV:A	5EHW:A	5EKH:A
5EKJ:A	5EKM:A	5FAI:A	5FDC:A	5FDI:A	5FLO:A	5FLP:A	5FLQ:A	5FLR:A
5FLS:A	5FLT:A	5FNG:A	5FNH:A	5FNI:A	5FNJ:A	5FNK:A	5FNM:A	5HWY:A

TABLE 2.1: Complete list of knotted proteins. The nomenclature is the following: PDB name:protein chain(model).

We have then computed for both open (after closure) and cyclized candidates the Alexander determinant to characterize the chain's topological state, and localized the position of the knot using the bottom-up approach. All the aforementioned methods were characterized in Chapter 1.

Following this procedure, we found about 800 knotted protein chains, which are listed in Table 2.1.

E-F: Identify the representatives

The list of table 2.1 is partly redundant. This is because the PDB usually contains homologs entries, as well as multiple structures of the same proteins in different conditions of temperature, pH, and possibly alternative models for the same entry. Accordingly, we started by removing redundant proteins with exactly the same sequence.

We have then clustered the remaining sequences by performing pairwise sequence alignments, in order to identify a minimally-redundant set of representative knotted proteins. The degree of sequence alignment between two sequences was assessed using the BLAST score [40], while the clustering was performed using the cd-hit suite [41], which identifies as representative for a given cluster its longest sequence. In particular, we have grouped together proteins with a sequence identity larger than a 10% threshold.

G: Elimination of artifactual knots

The set of representatives was further analyzed for artifactual knots that can possibly arise due to a limited experimental resolution in structure determination.

Indeed, based on detail inspection of the knotted region and its characteristics, we have identified various putative artifactual proteins listed in Table 2.2, for reasons discussed below.

PDB code	Method	Knot type	Knot length (a.a.)	Chain length (a.a.)	N-terminal knot depth	C-terminal knot depth
4V4B:I	El.Micr.	3_1	9	139	93	37
4V1A:U	El.Micr.	3_1	63	387	186	138
5AJ4:IC	El.Micr.	3_1	63	387	186	138
2LXD:A(3)	NMR	3_1	50	123	58	15
1J2O:A(14)	NMR	4_1	52	114	42	20
2KC2:A(2)	NMR	3_1	68	128	56	4
2KQU:A(20)	NMR	3_1	42	169	53	74
2KN4:A(2)	NMR	4_1	80	158	71	7
2Q7Z:A(5)	Sol.Scat.	3_1	123	1931	1225	583
3GAU:A(2)	Sol.Scat.	3_1	720	1213	138	355
3GAW:A(2)	Sol.Scat.	3_1	354	1213	426	433
3GAV:A(8)	Sol.Scat.	4_1	479	1213	319	415
3GAU:A(7)	Sol.Scat.	7_7	385	1213	403	425
3GAV:A(2)	Sol.Scat.	7_7	393	1213	396	424
3GAV:A(7)	Sol.Scat.	8_{18}	476	1213	518	219

TABLE 2.2: List of possibly artifactual protein representatives, bolded, along with their clustered homologs and ordered by resolving method and knot complexity. Note the presence of unlikely very complex knot topologies in the X-ray solution scattering case. For the PDB nomenclature, see Table 2.1.
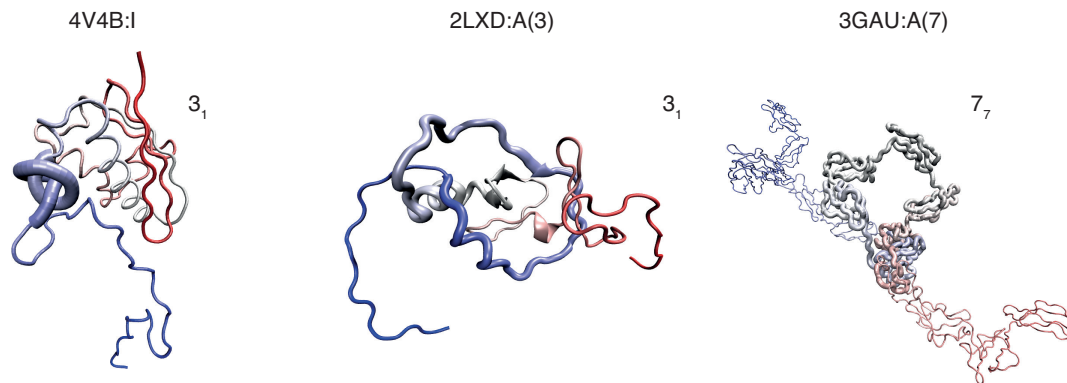
FIGURE 2.5: Three examples of artifactual proteins, taken from Table 2.2. The first example is a protein chain belonging to the *Saccharomyces cerevisiae* ribosome complex (PDB code 4V4B:I), resolved with cryo-EM, that contains a very narrow knotted core. The second example is a regulatory protein (PDB code 2LXD:A(3)) from solution NMR. The third example is a complement factor H resolved structure (PDB code 3GAU:A(7)) from X-ray solution scattering, with a complex topology.

The first three instances presented in Table 2.2 were resolved with cryo-EM. The first one is a protein chain of the *Saccharomyces cerevisiae* ribosome complex (PDB code 4V4B:I). The protein structure, shown in Fig. 2.5, presents an unrealistically tight knot of only 9a.a. that violates the steric hindrance of the protein residues. Moreover, it can be noted that a single clasp passage between two near strand can unknot the configuration, suggesting that the knot could have been formed by two near strands wrongly resolved. The second protein of Table 2.2, a mitoribosomal protein structure (PDB code 5AJ4:IC), was excluded along its clustered homolog (third in the list), as we cannot safely assess for the same reasons if the knot is genuine or artifactual.

The second group of instances presented in Table 2.2 were resolved with solution NMR. This method computationally reconstructs the protein structure based on several distance, angle and orientation restraints between atoms, and provides an ensemble of probable conformations (models). These instances were selected because there is not a consensus on the topology across the set of alternative models provided for the same structure, and in particular only some of them are knotted. We can also observe that in every case there are unstructured parts, see Fig. 2.5 for an example. Thus, these knotted proteins might be either artifactual or part of a wider configurational ensemble, e.g. of disordered proteins which genuinely could include both knotted and unknotted configurations. For this inherent ambiguity we omitted them from further analysis.

The last group of Table 2.2 is resolved with X-ray solution scattering. This method reconstructs the protein structure from an X-ray scattering pattern in solution, with the reconstruction of the three-dimensional structure providing several possible models. Similarly to solution NMR, only some of the models are knotted, thus creating the same ambiguity discussed above. Interestingly, in this case much more complicated topologies can appear, see Fig. 2.5.

Representatives

The minimally-redundant representative set, excluding the possible artifactual knotted proteins, is shown in Table 2.3. This table is built by ranking the representatives in order of topological complexity, then in order of functionality, and finally with the number of represented structures.

All representatives reported here are well documented from previous surveys [17, 37, 42, 43].

2.3   Knotted protein characteristics

From the analysis of the entire set of individual peptide chains in the PDB, several facts emerge.

The first is that the number of knotted chains in the PDB (about 800) is much smaller than the total number of chains (about $5 \cdot 10^5$). Keeping in mind that both sets are redundant, the ratio between these two numbers provides an estimate of the proteins knotting probability which is of the order of $800/5 \cdot 10^5 \sim 1\%$.

It is of particular interest to compare this number, which is small but still non-negligible, to simpler models reproducing the same globular characteristics of proteins.

In this regard, Lua and Grosberg [44] compared the protein's knotting probability to that of self-avoiding random walks on a lattice. In particular, the authors constructed walks with the same average end-to-end distance of proteins, and found that these walks had a much larger knotting probability than proteins. Next, the authors concluded that the difference in knotting probability could be ascribed to a different sub-chain structural organization: segments of the proteins with intermediate lengths have a stronger statistical tendency to fold back on themselves than those of the self-avoiding random walks. This suggests that in proteins the interpenetration of sub-chains is suppressed and, consequently, knot formation too.

These results are compatible with the intuitive expectations that knots have been selected against in naturally occurring proteins, to promote rapid folding, though not ruled out entirely.

The second property emerges by comparing the knot length of each knotted protein l_k to the total chain length N. This comparison provides an insight into the polymeric nature of a protein. The plot in Fig. 2.6 of l_k versus N, using only data from the representatives (in order to have an unbiased set) shows a significant correlation between l_k and N, in agreement with the $l_k \propto N^\alpha$ law of general flexible chains [45, 46]. In particular, the fitted $\alpha = 0.86$ can be compared with the values $\alpha = 0.63$ and 1.04 for a polyethylene chain model up to $N = 1000$ in coil (good solvent) and globular (bad solvent) form [45], and with $\alpha = 0.44$ for a freely-jointed chain up to $N = 15000$ [46]. We thus conclude that, although proteins have a different structural organization than general random polymer models, they too feature a concomitant increase of l_k with N.

The third property is obtained by measuring which terminal is closer to the knotted region, again considering only the representative proteins. Interestingly, we did not find any overwhelming preference over one of the two termini, as shown in Fig. 2.6. Because knots are formed by threading one terminus at least once over a loop, both termini can be involved in the knot formation, if we suppose the nearest terminus is the one responsible for the threading process. This result, in return, runs against the expectations that the C-terminus should be privileged, as the proteins fold *in-vivo* co-translationally starting from this terminus.
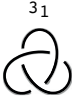
Knot type	Protein or protein function	PDB code	Oligomeric state	Knot length (a.a.)	Chain length (a.a.)	N-terminal knot depth (a.a.)	C-terminal knot depth (a.a.)
3_1	Carbonic anhydrase	3MDZ:A	Monomeric	230	259	26	3
	RNA methyltransferase	1X7O:A	Homodimeric	44	267	190	33
	RNA methyltransferase	4H3Z:B	Homodimeric	49	256	89	118
	RNA methyltransferase	4CND:A	Homodimeric	57	169	79	33
	RNA methyltransferase	4E8B:A	Homodimeric	49	242	164	29
	RNA methyltransferase	3O7B:A	Homodimeric	47	216	143	26
	RNA methyltransferase	4JAK:A	Homodimeric	44	155	77	34
	RNA methyltransferase	4JWF:A	Monomeric	46	187	98	43
	RNA methyltransferase	2QMM:A	Homodimeric	45	195	124	26
	S-adenosylmethionine synthetase	4ODJ:A	Homodimeric	258	386	16	112
	Carbamoyltransferase	3KZK:A	Homodimeric	83	334	169	82
	Hypothetical RNA methyltransferase	1O6D:A	Homodimeric	50	147	67	30
	Hypothetical protein MJ0366	2EFV:A	Homodimeric	63	82	10	9
	H+/Ca2+ exchanger	4KPP:A	Monomeric	218	395	72	105
	Na+/Ca2+ exchanger	5HWY:A	Monomeric	197	300	33	70
	N-acetylglucosamine deacetylase	5BU6:A	Monomeric	249	264	10	5
	DNA binding protein	2RH3:A	Monomeric	102	121	7	12
	DNA binding protein	4LRV:A	Homo-octameric	88	107	8	11
	Metal binding protein (zinc-finger)	2K0A:A	Spliceosome component	52	109	21	36
4_1	Bacteriophytochrome	4GW9:A	Homodimeric	243	628	20	365
	Ketol-acid reductoisomerase	1QMG:A	Homodimeric	210	514	236	68
5_2	Ubiquitin carboxy-terminal hydrolase	2LEN:A	Monomeric or homodimeric	217	231	2	12
6_1	α-Haloacid dehalogenase I	3BJX:B	Homodimeric	216	295	59	20

TABLE 2.3: Representative knotted proteins. This up-to-date, non-redundant list of knotted representatives is based on a PDB survey specifically carried out for our review [38]. The entries are presented in order of increasing complexity of the knot and, for each knot type, they are listed in order of decreasing number of represented homologs (number of PDB entries with sequence identity larger than 10%). The oligomeric state is based on the PDB and UNIPROT annotations. For a stringent determination of physical knots, chains with structural gaps, or with termini not exposed at the protein surface were excluded from survey. For the PDB nomenclature, see Table 2.1.
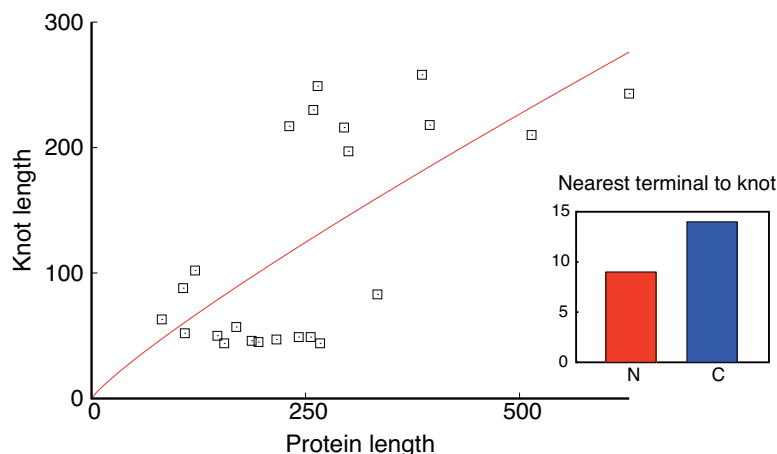
FIGURE 2.6: Length and depth of knots in proteins. The scatter plot presents the knot length versus protein length for the 23 minimally redundant representatives of Table 2.3. The Kendall's correlation coefficient, $\tau = 0.31$ and the one-sided p-value is 0.018. The red curve is a best fit of the $l_k = aN^\alpha$ function, with $a = 1$ and $\alpha = 0.86$. Of the 23 instances, 9 have the knotted region closest to the N terminus and 14 to the C one, see inset.

The last aspect emerges from the different knot types found in Table 2.3, which correspond to the simplest instances of twist knots ($3_1, 4_1, 5_2$ and 6_1). The preference for this knot family might be related to the fact that different knot types require a different minimum number of threading processes [42]: the twist knot family require only one threading passage (see Sec. 1.2), while torus knots for example require in general more passages, aside from the 3_1 (which is also a twist knot). However, other knot types, not presently observed, cannot be ruled out entirely, because the PDB is still growing adding new resolved structures everyday.

2.4   Details of notable knotted protein case studies

I will review here some notable cases of knotted proteins present in Table 2.1.

Carbamoyltransferases
The representative proteins in Table 2.3 contain a Carbamoyltransferase protein, which is an enzyme with a 3_1-knot, see Fig. 2.7a. These kind of enzymes are important because they have unknotted homologs, and are an evolutionary sub-branch of the latter. Thus, knotted and unknotted homologs can be compared structurally, to see the actual differences and if there is some evolutionary pattern that brought forth a knotted fold.

In this regard, Virnau *et al.* [17] observed that the knotted Carbamoyltransferase have a rigid loop through which the knot is threaded, and which encompasses the active site of the protein, possibly altering the enzyme's catalytic activity. Afterwards, Potestio *et al.* [37] observed that there are two short loops inside the knotted Carbamoyltransferase, whose virtual excision eliminates the knot altogether, see Fig. 2.7b, and that this aspect is common to other knotted proteins with unknotted homologs, concluding that these loops might promote knot formation.

a) Carbamoyltransferase
3l04:A

b) Knotted and Unknotted Carbamoyltransferases
2fg6:C (knotted)   1ort:A (unknotted)
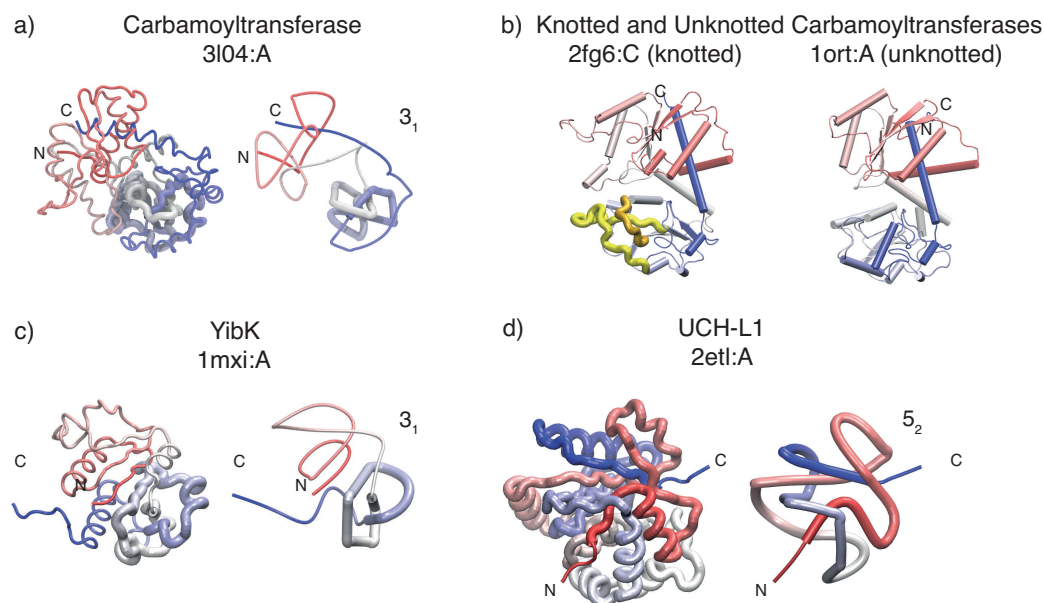
c) YibK
1mxi:A

d) UCH-L1
2etl:A

FIGURE 2.7: Selected examples of knotted proteins, from [38]. In panels (a, c, d) a smoothed structural representation is used to highlight the knot. Panel (b) presents a knotted/unknotted pair of carbamoyltransferases [37]. A virtual excision of either or both of the highlighted loops (coloured in yellow and orange) unties the knotted variant.

YibK and YbeA

Several representatives belong to the family of methyltransferases, and all are trefoil knots.

Most of the experimental investigations [47–51] of these knotted enzymes were carried out by the group of S. Jackson, who focussed on two of the smallest proteins of the methyltransferase family, YibK, shown in Fig. 2.7c, and YbeA. Jackson and coworkers were in particular interested on the robustness of knotted structure after denaturation, the role of chaperonine molecules in the aid of folding, and a possible characterization of the folding process.

Both YibK and YbeA were unfolded with high concentrations of chemical denaturant and cyclized thereafter in order to fix their topology. The denaturant was then removed and the protein refolded. The experiments showed that both proteins retained the same initial topological state 3_1, demonstrating that the knotted portion can be highly resilient to degradation.

In another experiment [51] an extra 91-residue thermophilic protein, *Archaeoglobus fulgidus* ThiS, was fused with each or both protein ends. The ThiS protein folds faster than YibK and YbeA, and into a compact globule, constituting a possible obstacle for the threading process. By trying all possible combination of ThiS attached only to one of the two ends or both, it was possible to investigate whether a preferred threading terminus exists. In particular, the folding rate was showed to slow down by up to a factor of three when the ThiS plug was attached to the C terminus. No appreciable slowing down was observed when the plug was present solely at the N terminus. These results suggested that the rate-determining folding and knotting events involve the C terminus, close to the location of the knotted region in the native state.

The same experiment [51] showed that chaperonines, which are rescue proteins, give also a dramatic increase in the folding efficiency of the knotted proteins. This evidence might hint at how knotted proteins might have withstood evolutionary pressure in order to have an efficient folding.

Complex knotted proteins: UCHs

The ubiquitin C-terminal hydrolase (UCH) family is characterized by a 5_2 topology, located only a few residues from the N terminus. Both UCH-L1, shown in Fig. 2.7d, and UCH-L3 have been probed with several techniques. UCH-L1, in particular, constitute the most abundant protein in the brain (2% ot the total number of brain's protein), and its mutation may cause a partial loss of its catalytic activity and may lead to protein's aggregation, associated with Parkinson's disease [52, 53]. Thus the knot, which spans the substrate binding and catalytic site, has been hypothesized to be instrumental in the protein avoiding proteasomal degradation [43].

Structural comparisons with unknotted homologs showed, similarly to Carbamoyltransferases, the presence of additional short loops compared to the unknotted cases [37].

It was experimentally shown that refolding after unfolding with chemical denaturants occurs via two parallel pathways, each with a metastable intermediate [54–56], neither of which was knotted, consistent with the rate-limiting step associated with the threading event required to establish eventually the 5_2 topology.

An hint about knotting causing difficulties in proteasomal degradation comes from simulations performed with a proteasome machinery represented by an effective potential with an added pulling force [57]. The presence of a knot is shown to hinder and sometimes even jams the process. It is likely that 5_2-knotted proteins such as UCH-L1, which have much larger twist knots in their denatured states will have an even greater effect, as observed in the systematic pore translocation computational study of Ref. [58], which will be discussed in Chapter 4.

Experiments showed also that mutants associated with an increased risk of Parkinson's disease, might be subtly related to different internal arrangement of the side chains, with exposure of hydrophobic surface area with an associated increase in *in-vivo* aggregation and other *in-vivo* aberrant interactions [54].

2.5   Conclusions

Protein native folds were long believed to be knot free, because the formation of knots during the folding process might introduce kinetic traps that could slow down folding. Despite this belief, several knotted proteins with increasing complexity were found, starting from the first systematic survey of 1994 [9].

The ever increasing size of the PDB motivated us to carry out an up-to-date survey of the knotted entries present in the PDB. We found 800 knotted entries out of $5 \cdot 10^5$ processed peptide chains. After removing sequence redundancy, the resulting knotted protein families were about 23. The survey highlighted the presence of several knotted proteins that are likely artifactual products of limited experimental resolutions, and therefore excluded from the analysis.

Afterwards, I considered several characteristics of knotted proteins. We found a knotting probability for proteins of about 1%, consistent with previous studies. Such probability is intriguingly smaller than the expected theoretical one based on models that reproduce the structural salient properties of proteins. In particular, it was

argued that the difference could arise from the different sub-chain structural organi-zation of proteins compared to that of generic self-avoiding walks, which might rule out as much as possible knotted conformations. However, proteins still preserve a degree of knot localization compatible with that of general polymer chains, because there is a significant correlation between the knot length and the entire chain length considering the knotted representatives.

We also inferred properties regarding the kinetics of knot formation, which in-volves the passage of at least one terminal through a loop formed by the protein chain. In particular, there does not seem to be a preferred terminus nearer to the knot, and this probably implies that both termini can be equally involved in the threading process depending on the specificity of the protein sequence. At the same time, we also have observed that the current knotted topologies present in the PDB are remarkably all twist knots. Their common aspect is that they can all be formed with a strand passage, contrary to other knot types not present.

We concluded by discussing notable example of knotted entries, for which ex-tensive experiments or numerical results are available.

Chapter 3

Knotting dynamics of DNA chains confined in nanochannels

In this chapter, I will characterize, by molecular dynamics simulations, the typical mechanisms governing the spontaneous tying, untying and the dynamical evolution of knots for model DNA chains confined in nanochannels.

The chapter is organized as follows: first I will summarise the established phenomenology of DNA knotting in various contexts and motivate the relevance of extending considerations to DNA chains inside nanochannels. Next, after introducing our coarse-grained model of choice for DNA, I will present a detailed analysis of the Langevin dynamics simulations and discuss how the interplay of DNA length and chain confinement affects the kinetics of formation and untying of knots and their lifetimes.

The content of this chapter is based on the work published in Ref. [59].

3.1 Introduction

3.1.1 DNA and knots

After the celebrated work of Watson, Crick, Franklin and Wilkins, DNA has been commonly regarded as the polymer of life *par excellence* because the genetic information of any organism is encoded in the chemical composition of its DNA. DNA is typically found in the double-stranded form, resulting from the intertwining of two chemically-complementary filaments, each consisting of a sequence of four different types of nucleotides (adenine, thymine, guanine and cytosine or A T G C).

In all organisms, DNA is subject to a significant degree of spatial compactification because it is confined in regions whose transverse size is much smaller than the DNA contour length.

The total contour length of DNA filaments (chromosomes) in eukaryotic cells is in the range of $[10^{-3} - 1]$m (corresponding to $[10^7 - 10^{11}]$ base pairs (bps), and is all kept inside the nucleus whose diameter is only about $\sim 6\mu$m.

In prokaryotic cells such as bacteria, the genome is typically $[0.034 - 3.4]$mm-long (corresponding to $[10^5 - 10^7]$bp) and is packed in a region of transverse size of about $\sim \mu$m.

Finally, bacteriophages DNA, which is several microns long, is kept inside viral capsids that have a diameter of about 50-80nm, which is comparable to the persistence length of DNA. We recall that latter is the arclength over which the directional correlation of DNA persists. More precisely, it is the charachteristic decay arclength of the tangent-tangent correlation. At 300K, the persistence length of ds-DNA is about $l_p = 50nm$, or about 150 bps.

Considering the degree of spatial compactification of the genome inside all organisms, it appears inevitable that DNA filaments *in vivo* must experience significant amounts of entanglement and knotting [11–13, 15, 60–64].

Indeed, a considerable amount of experimental evidences for DNA knotting has been gathered in the past decades, especially for viral DNA. The genome of viruses, in fact, is short enough that its topological state can be assessed directly, with electron microscopy, or indirectly using gel electrophoresis. The latter technique, in particular, can separate different knot types in chains with the same length: knots migrate along the gel with different velocities according to their topology, because they engage with and negotiate the gel obstacles in different manners, see Fig. 3.1.
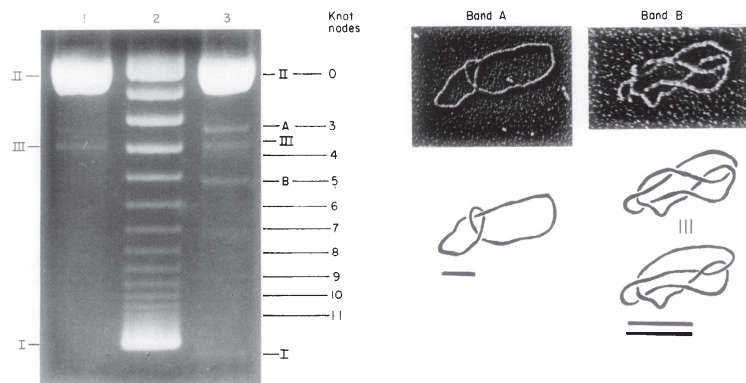

FIGURE 3.1: On the left, separation of the knotted DNA species using agarose gel electrophoresis in different bands related to the number of crossings. On the right, an electron microscopy image of knotted DNA molecules with 3 and 5 nodes, recovered from band A and B and coated with RecA protein in order to enlarge their radius size. From [65].

Knots were first directly observed on a circular phage DNA (composed by a single strand) in 1976 [11]. Thereafter, the abundance of knots was quantitatively measured by performing random circularization on linear P4 DNA [12, 60] and in plasmids purified directly from *Escherichia coli* bacteria [65].

In the case of DNA-viruses, which are much more simpler than bacteria or Eukaryotic cells, some insight about the organization process was given by studying *in-vivo* abundance of DNA knots [14, 61, 66, 67] and combining it with molecular dynamic simulations [68, 69].

3.1.2 Knotting of DNA inside nanochannels

In recent years, there has been much interest in characterizing the behaviour and physical properties of DNA under spatial confinement [70–76]. The intrinsic constraints of DNA, such as the persistence length, contour length and excluded volume, in relation to the geometric constraints of the environment, such as its size and dimensionality, may greatly influence the metric, kinetic and topological properties of the chain. The confining environment can be of different kinds: sphere [14], slit [77, 78] or channels [70].

DNA confinement inside nano-channels offers in particular a test-ground for polymer's theory, because it creates a competition between different length-scales that are both intrinsic to the chain and extrinsic. The former include the chain contour length L_c, the gyration radius R_g, which measures the polymer's size and the

persistence length l_p. The extrinsic length in this case is the channel diameter D, imposed on the chain. Such competition between lengths, in turn, produces various metric scaling regimes for the longitudinal chain extension.

Two limiting scaling regimes have been particularly studied, see Fig. 3.2.

For $D >> l_p$ and D smaller than the gyration radius of the chain in bulk, the polymer is in the so-called de Gennes scaling regime [79], developing a series of consecutive blobs that repel each other, see Fig. 3.2a. The longitudinal chain span scales as $L_c(D/l_p)^{-0.70}$.

If D is, instead, reduced well below l_p, the back-folding is energetically disfavoured, and the chain elongates as much as possibile, with a series of deflections from the wall, see Fig. 3.2b. In this case, the span scales as $L_c(1 - A(D/l_p)^{2/3})$ and we are in the Odijk deflection regime [80].

These two limiting regimes are bridged by a crossover region, where further different scaling mechanisms have been argued to exist [74, 81–86].

Besides considering metric properties, relevant insight on channel-confined DNA can be obtained by profiling its self-entanglement, and particularly knotting [87–90]. This is both interesting from the fundamental point of view of polymer theory, as well as for its practical implications, because knots can potentially interfere and limit the scope of nano-manipulation techniques based on nanochannels, such as DNA length sorting or DNA fingerprinting using barcoding techniques [91–94].
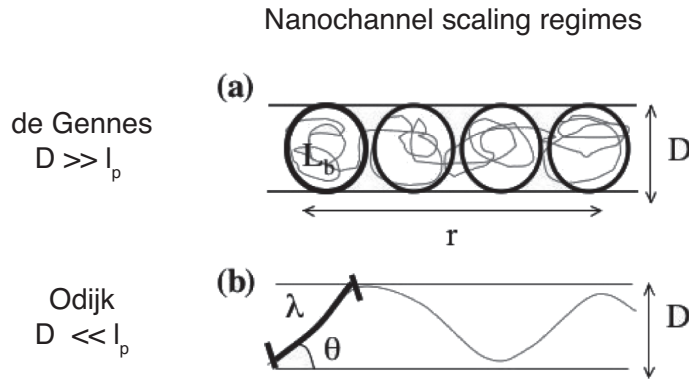
<div align="center">Nanochannel scaling regimes</div>

FIGURE 3.2: Different scaling regimes of a polymer of persistence length l_p confined inside a nanochannel. a) De Gennes regime, with $D >> l_p$. The polymer is subdivided in a series of blob of size L_b that repel each other. b) Odijk regime, with $D << l_p$. The polymer has an elongation that depends on the number of deflections made by the polymer with the walls, with $\lambda = (D^2 l_p)^{\frac{1}{3}}$, and with back-folding energetically unfavored. Image from [76].

In this regard, we expect that the bulk knotting probability of DNA should drop to zero in the Odijk regime, because back-folding and so the threading process will be suppressed altogether. However, Micheletti and Orlandini observed, with Monte Carlo simulations [87], that the knotting probability is a non monotonic function varying the channel width, see Fig 3.3. In particular, for a given chain length, p_{knot} was found to first increase steadily from its bulk value as D was reduced, reaching its maximum value for $D \sim 80$nm, which approximately corresponds to the onset of the crossover to the Odijk regime. Upon further reducing D below 80nm, an abrupt drop of the knotting probability was finally observed.

The insight offered by this equilibrium study was complemented by a study of the dynamical mechanisms governing the spontaneous appearance and untying of
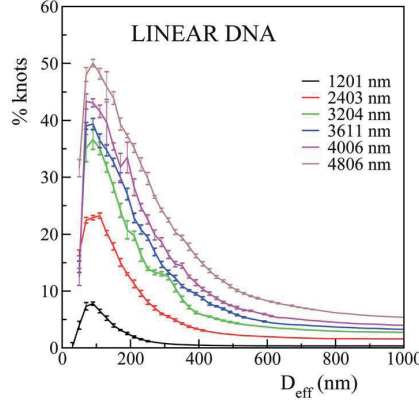
FIGURE 3.3: Equilibrium knotting probability of DNA as a function of channel width D and chain length, from [87].

knots in DNA chain inside nano-channels [95]. In this study the chain length $L_c = 4.8\mu$m was kept fixed and the channel width D varied, clarifying that by decreasing D knots were suppressed due to a a shallower and more infrequent back-folding process.

The latter study [95] did not clarified, however, the dynamical mechanisms responsible for the increase in the knotting probability increasing the chain length L_c.

To address this open question, I will use molecular dynamic simulations to characterize chains of varying length, from 1.2 to 4.8μm, inside a channel of fixed width below the p_{knot} crossover region. In particular, I will study the knotting dynamics of the chain by evaluating the length dependence of the depth at which knots are formed or untied, as well as the lifetimes of unknotted and knotted states.

The content of the chapter is based on Ref. [59].

3.2 Methods

3.2.1 Model and simulation details

DNA is modelled as a self-avoiding chain of N beads of diameter equal to σ, mapped to a value of 2.5nm (equal to the hydration radius of double stranded DNA) and endowed with a bending rigidity set by the phenomenological persistence length, $l_p = 50$nm. The chain potential energy is accordingly given by three terms:

$$
\begin{aligned}
U_{\text{LJ}} &= \sum_{i,j>i}^{N} 4\epsilon \left[\left(\frac{\sigma}{d_{i,j}} \right)^{12} - \left(\frac{\sigma}{d_{i,j}} \right)^{6} + \frac{1}{4} \right] \theta(2^{1/6}\sigma - d_{i,j}) \\
U_{\text{FENE}} &= -\sum_{i}^{N-1} 15\epsilon \left(\frac{R_0}{\sigma} \right)^2 \ln \left[1 - \left(\frac{d_{i,i+1}}{R_0} \right)^2 \right] \\
U_{\text{bend}} &= \sum_{i=2}^{N-1} \epsilon \left(\frac{l_p}{\sigma} \right) \left(1 - \frac{\vec{b}_{i-1} \cdot \vec{b}_i}{\left| \vec{b}_{i-1} \right| \left| \vec{b}_i \right|} \right).
\end{aligned}
\tag{3.1}
$$

In the above expression, $d_{i,j} = |\vec{r}_i - \vec{r}_j|$ represents the distance between two beads centers i and j, $\vec{b}_i \equiv \vec{r}_{i+1} - \vec{r}_i$ is the vector bond connecting two consecutive beads of the chain and θ is the Heaviside function ($\theta(x) = 0$ if $x < 0$ and $\theta(x) = 1$ if $x \geq 0$).

ϵ is the characteristic unit of energy of the system, set equal to the thermal energy, $k_B T$.

U_{LJ} is a truncated and shifted Lennard-Jones potential accomplishing the excluded volume interaction between any pair of beads, including consecutive ones: two beads start to repel if their distance is less than $2^{1/6}\sigma$, which correspond to the minimum of the potential.

The FENE potential U_{FENE} [96] is a spring potential that enforces the connectivity of the chain, so that two consecutive particles cannot distance more than $R_0 = 1.5\sigma$.

U_{bend} is the bending energy, which penalizing consecutive bond vectors that are not parallel. The latter in particular enforces that the persistence length is the right one at a given temperature.

In addition the chain is subject to the excluded volume interaction with the walls of the confining channel of width D, and axis corresponding to the x Cartesian axis. This interaction is described with a truncated and shifted Lennard-Jones potential too:

$$U_{wall} = \sum_i^N 4\epsilon \left[\left(\frac{\sigma}{\delta_i}\right)^{12} - \left(\frac{\sigma}{\delta_i}\right)^6 + \frac{1}{4} \right] \theta(2^{1/6}\sigma - \delta_i), \qquad (3.2)$$

where $\delta_i \equiv \frac{D}{2} - \Delta_i$ and Δ_i is the distance of the ith bead from the channel axis.

The dynamical evolution of the chain (initially prepared in an perfectly straight conformation lying along the channel axis), is described with a Langevin dynamics:

$$m\ddot{\vec{r}}_i = -\gamma\dot{\vec{r}}_i - \nabla U + \vec{\eta} \qquad (3.3)$$

where γ is the friction and $\vec{\eta}$ is the noise, a Gaussian random variable with zero mean and delta-correlated variance. The variance of each component of the noise $\sigma_\eta^2 = 2\gamma k_B T$ satisfies the fluctuation-dissipation relation. We also choose $m/\gamma = 2\tau_{LJ}$, with $\tau_{LJ} = \sigma\sqrt{m/\epsilon} = \sigma\sqrt{m/k_B T}$ being the characteristic simulation time (see [96]).

We considered individual chains of N=480, 960, 1437 and 1922 beads (or equivalently of contour length $L_c = 1.2, 2.4, 3.6, 4.8\mu$m) confined inside channels of width $D = 56$nm.

The dynamical equation is integrated numerically using the LAMMPS simulation package [97], setting as customary $k_B T = \epsilon = 1$, $m = 1$ and $\sigma = 1$. We chose as integration timestep $0.012\tau_{LJ}$, sampling the chain configuration every 10^5 timesteps.

For each chain length we collected about 10 independent trajectory, covering a total of 10^{11} integration time steps. For the case $N = 1437$ the analysis included 3 trajectories for a total of 3×10^{10} integration time steps that were originally collected for the publication of ref. [95] . All other simulations were specifically carried out for this study.

We can map time from simulation units to physical units considering the self-diffusion coefficient of an isolated bead $D = \frac{k_B T}{\gamma}$ and matching it to $D_{stokes} = \frac{k_B T}{6\pi\eta\sigma/2}$, estimated by applying Stoke's law to a bead of diameter $\sigma = 2.5$nm immersed in a fluid of viscosity equal to water, $\eta = 1cP$. By using this nominal mapping (which neglects hydrodynamic and wall-proximity effects), and assuming $T = 300K$, one obtains that $\tau_{LJ} = \frac{6\pi\eta\sigma^3}{\epsilon} = 74$ns, or equivalently a single timestep $0.012\tau_{LJ} \sim 1$ns. Throughout the manuscripts we shall use this mapping to provide time in physical units.

3.2.2 Knotting properties: type and location

To establish the knotted state of a given configuration, we first closed the open chain with an auxiliary arc selected using the minimally-interfering closing procedure and then detected the knot type with the Alexander determinant. These methods are described in Chapter 1. In this way we assign a definite topological state to each configuration sampled in the MD trajectory.

We enhanced the robustness of the identified topological state by establishing it through a majority rule applied to 21 consecutive sampled conformations (each sampled every 100μs), following Ref. [95]. In this way, we removed ephemeral changes of topology lasting for less than 1ms, that might be due to slight repositionings of the chain ends during the trajectory, or to the uncertainty in the chain closure procedure if one or both ends is buried, as occurred in proteins.

We also identified the portion of the chain that accommodates the physical knot with the bottom-up approach. In this way we can infer the starting and ending bead of the knot, along with its average position along the chain and the contour length it occupies. In order to characterize in an absolute way the depth at which the knot is present, we defined it as the shortest contour length between its average position and one of the two terminal beads of the chain. So this value can range between 0 and $L_c/2$.

3.3 Results and discussion

3.3.1 Knotting probability

We analysed the dynamical evolution of knots in DNA chains of length $L_c = 1.2$, 2.4, 3.6 and 4.8μm individually confined inside a channel of width $D = 56$nm. This width is comparable with the DNA persistence length, $l_p = 50$nm and hence lies at the crossover between the de Gennes and Odijk metric scaling regimes, where chain back-foldings are disfavored but not completely suppressed. Typical knotted and unknotted configurations of the confined DNA are shown in Fig. 3.4.
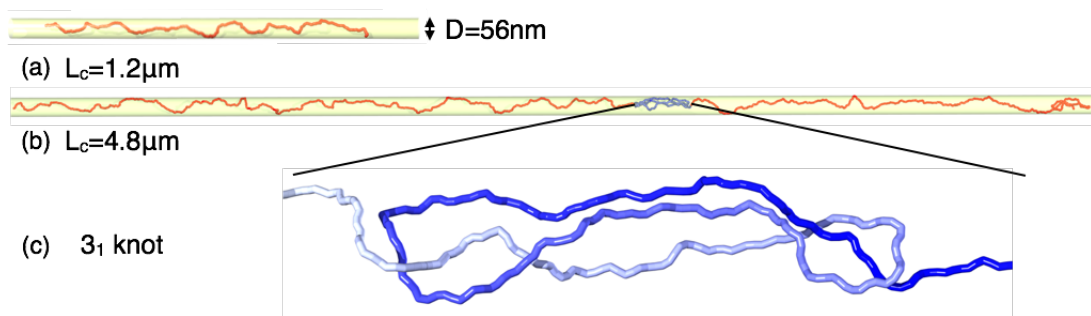

FIGURE 3.4: Typical equilibrium configuration of the shortest (a) and longest (b) chain considered in our study ($L_c = 1.2, 4.8\mu$m) inside a channel of width $D = 56$nm. The configuration in panel (b) accommodates the trefoil, or 3_1-knot shown in panel (c).

In order to have an estimate of the knotting probability that is unbiased with respect to the initial, unknotted, configuration we discard the set of samples preceding the first knotting event; this ensures that the overall sampling is topologically equilibrated and the unknotted events are independent on the initial configurations.

The average knotting probability, p_{knot}, was computed by computing the number of knotted configuration and dividing it by the total number of configurations, sampled across all trajectories and at each value of L_c. The resulting probability profile is shown in Fig. 3.5a and displays a marked growth from 2% to 13% upon increasing the chain length from 1.2μm to 4.8μm. This is qualitatively understood with the intuitive argument that, at fixed D, longer chains can accommodate more knots than shorter ones.

Although the limitations of the few dataset points should be borne in mind, p_{knot} appears to have a noticeable non-linear dependence on L_c over the considered range of chain lengths. In fact, its best-fit power law, $p_{knot} \sim aL_c^b$, shown in Fig. 3.5a, is associated to the exponent $b = 1.5 \pm 0.2$. At all considered chain lengths, the most abundant knot type is the 3_1, or trefoil knot, which represents between 76% (for the longest chains) and 88% (for the shortest ones) of the knotted configurations.
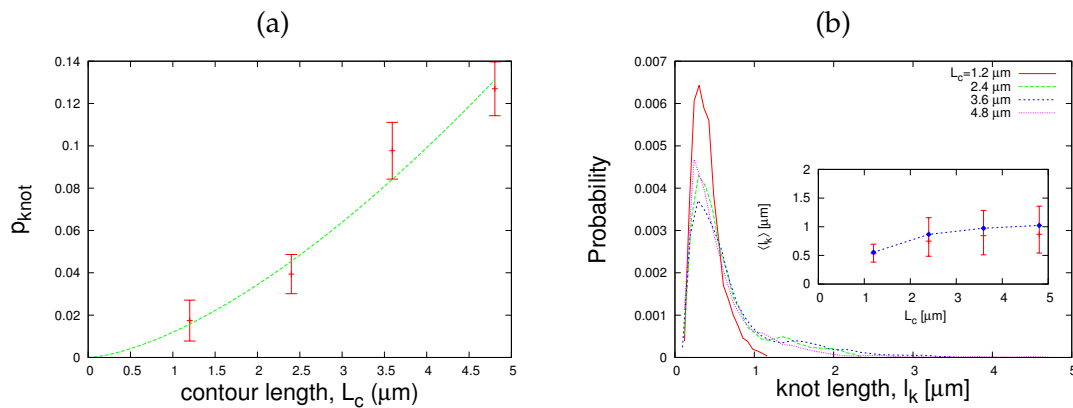

FIGURE 3.5: (a) Knotting probability, p_{knot}, as a function of the contour length, L_c, of the channel-confined chains. The dashed (green) curve is the best-fit power law, aL_c^b, where $a = 0.012 \pm 0.004$ and $b = 1.52 \pm 0.24$. The associated probability distributions of the knot length l_k are shown in panel (b). The inset shows the contour length dependence of various statistical properties of l_k such as its mean value (blue diamond) the median one (central red dash) and first and second quartile (top and bottom red dashes). Notice that $\langle l_k \rangle$ levels off as L_c is increased.

A further relevant equilibrium property is the average knot length, $\langle l_k \rangle$, which is shown in Fig. 3.5b. It is noteworthy that, at the considered channel width D, $\langle l_k \rangle$ does not vary significantly across the various chain lengths. In fact, even the normalised probability distribution of l_k is well consistent at $L_c = 2.4, 3.6, 4.8\mu$m. This represents a major qualitative difference with respect to the bulk (unconfined) case where $\langle l_k \rangle$ is known to increase appreciably with L_c following a sublinear trend, $\langle l_k \rangle \sim L_c^t$ with $t \sim 0.70$ [32, 98, 99].

It is worth pointing out that the average knot length, which is about 1μm represents a significant fraction of the contour length of the shortest chain.

3.3.2 Lifetimes of knotted and unknotted states

To clarify the relationship of chain knotting equilibrium properties with the dynamics inside the channel, we analysed several observables.

As a first step we analysed the length dependence of the duration of knotted and unknotted states, which will be denoted with τ_k and τ_u, respectively. These two quantities are, in fact, directly related to the knotting probability through $p_{knot} = \langle\tau_k\rangle/(\langle\tau_k\rangle + \langle\tau_u\rangle)$, where $\langle\cdots\rangle$ denotes the average of different knotted or unknotted intervals.
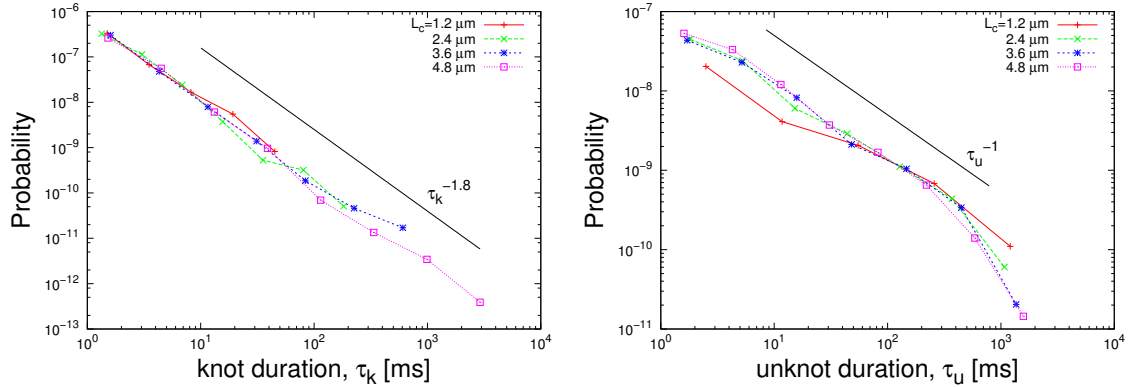

FIGURE 3.6: Probability distribution of the time duration of knotted (a) and unknotted (b) events. (a) For knotted states note that, as L_c increases, there are events that contribute to extend the tail of the distribution (long lived events) which decays approximately as $\tau_k^{-1.8}$, as indicated by the black solid line. This decay was obtained by fitting the data points for $\tau_k > 10$ms with a power-law. By contrast, the longest duration of unknots does not significantly depend on L_c, see panel (b), and the probability distribution decays approximately as τ_u^{-1}, as indicated by the black solid line.

It is seen from Fig. 3.6 that the probability distributions of both knots and unknots lifetimes have a monotonically decreasing trend that is largely consistent at the various chain lengths (particularly excluding the case $Lc = 1.2\mu$m where the chain length closely competes with the average knot length).

It is noteworthy that the decay of the knots curve is well described by a power law, $\tau_k^{-1.8}$. This, in turn, implies that, for any given L_c, the average knot duration is dominated by the relatively rare long-lived knots rather than by the frequent short-lived ones. At a general level, this observation crucially highlights the necessity to collect extensive dynamical trajectories for a reliable characterization of the salient knotting properties.

More specifically, the fact that in Fig. 3.6a the longest knot durations, i.e. the curves endpoints, increase very rapidly with L_c implies that the average knots lifetimes should grow significantly with L_c. By contrast, no appreciable length dependence is observed for the lifetime of unknots, see Fig. 3.6b, so that the average duration of unknots is expected to vary only modestly with L_c.

These expectations are, in fact, supported by the direct calculation of the average knots and unknots durations, which are shown in Fig. 3.7. As L_C increases, the lifetime of unknots gradually levels off with a decreasing trend, while the lifetime of knots keeps increasing.

It should be noted that the growth of the average knots duration is paralleled by the increase of its estimated error. This is consistent with the observation made above that the average knots lifetime is mostly controlled by the longest-lived knots which, as L_c increases, require longer and longer simulations to be reliably sampled.
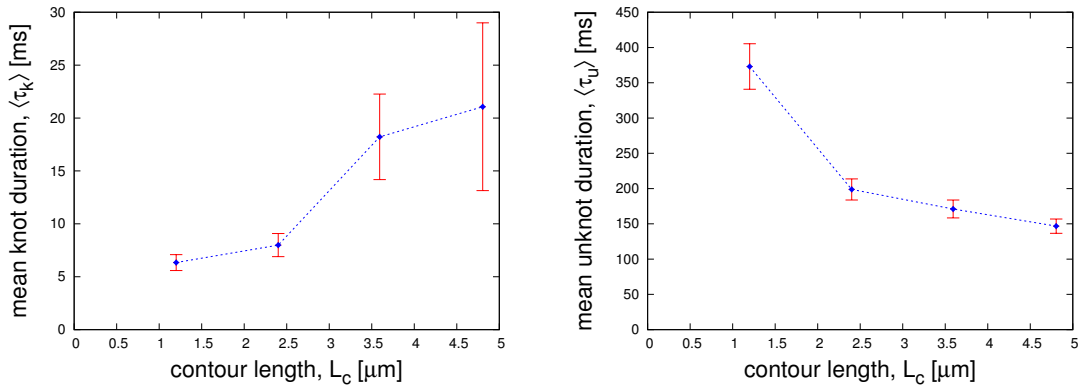
FIGURE 3.7: Mean duration of knots (a) and unknots (b) as a function of the contour length L_c.

3.3.3 Tying and untying dynamics of knots

The data presented so far aptly complement the equilibrium properties shown in Fig. 3.5a (average knotting probability), by showing that, at a given (moderate) level of channel confinement, the knotting probability increases with L_c mostly due to the increased lifetime of knots, while that of unknots varies to a lesser extent.

We can further clarify these properties by examining where, along the chain, knots typically dwell, or are tied and untied. Notice that the tying and untying event can be treated on equal footing because Langevin trajectories are statistically time-reversible.

The data shown in Fig. 3.8a clarify that most events leading to the spontaneous tying or untying of knots, occur close to the ends at a depth that is about independent on chain length. The shown data clarify that the same conclusion is reached by limiting considerations to knots with duration longer than $10ms$ (i.e. by excluding the short-lived knots that, though frequent, do not contribute appreciably to the average knotting probability).

By contrast, the average depth at which knots are observed during their whole lifetime increases steadily with L_c, see Fig. 3.8b.

These two aspects can be seamlessly rationalised along with the recent results of ref. [95], which focused on the knotting dynamics of chains of fixed length $L_c = 3.6\mu$m but confined in channels of varying width. The latter study of Micheletti and Orlandini clarified the typical knotting mechanisms occurring at confinements within the de Gennes-Odijk metric crossover region. These involve the formation of a simple knot via back-folding at one of the chain ends and its subsequent stochastic motion along the chain contour until it is eventually untied at one of the ends, again via chain back-folding (possibly after accumulating and releasing additional entanglement).

This observation, in fact implies that the lifetime of unknots, being the typical waiting time between two knotting events, should depend on the rate at which significant back-foldings of the ends are created. As this process is mostly localized near the chain ends, the duration of unknots should be largely independent on L_c. The lifetime of knots, instead, is largely controlled by their stochastic motion along the contour of the chains that are elongated inside the confining channel. As a matter of fact, the behaviour is similar to the observed diffusive-like motion of knots in mechanically-tensioned chains [100–105]. For this reason, knots that form in longer
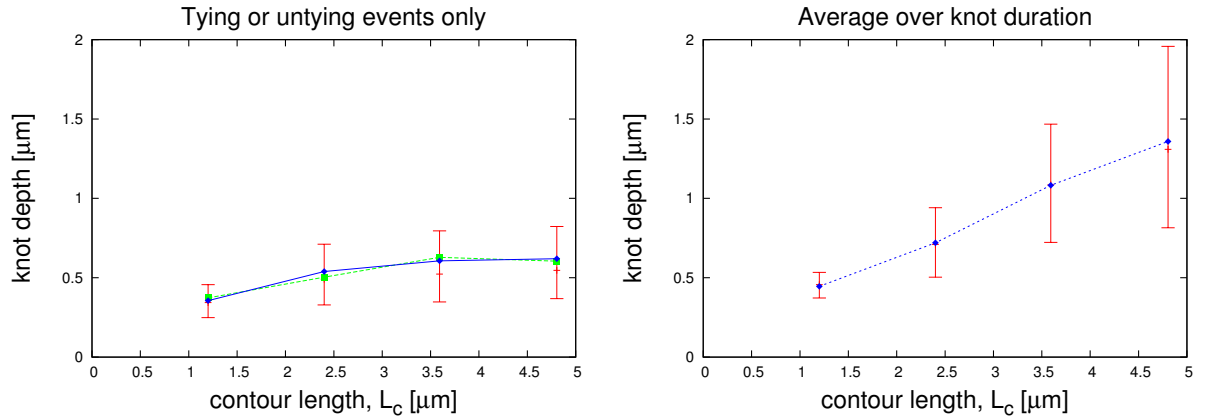
FIGURE 3.8: (a) Mean position at which knots are tied or untied (blue diamonds), along with the corresponding median and first and second quartile (red bars). The green squares indicate the mean depth at which knots of duration longer than 10ms, are tied or untied. The consistency with the mean depth computed over all knots suggests that, at the given level of confinement, the duration of knots does not significantly depend on the depth at which they are tied or untied. (b) Mean depth of knots (blue diamonds) averaged over their entire lifetime, i.e. averaged over all sampled knotted configurations. Again median, first and second quartile are shown as red bars. Notice the increase of the average knot depth with the chain length. The knot depth is calculated as the contour length separating the knot midpoint from the nearest chain end.

chains can, on average, reach more deeply the inside of the chain, and this, in turn, expectedly reflects in a significant growth of their lifetime with L_c.

An analogy can be drawn with a stochastic survival process where a particle, representing the knot, is created at a position x inside a segment, representing the elongated chain, covering the $[0, L_c]$ interval and then freely diffuses on it until it is absorbed by one of the segment ends. In such case, it can be shown that the average particle lifetime, τ, is proportional to $x(L_c - x)$ [106] and hence grows systematically with the segment length L_c. In particular, in the special cases where x is located in the middle of the chain or at a fixed small distance from a segment end, one has that τ depends, respectively, quadratically or linearly on L_c. Although the analogy holds only approximately (for instance because both the knot size and the chain configuration evolve with time) the argument gives a heuristic support to the more-than-linear increase of knots lifetimes portrayed in Fig. 3.7.

3.4   Conclusions

By using Langevin dynamics simulation we studied the spontaneous creation and subsequent evolution of knots in coarse-grained models of DNA chains confined in channels of fixed width $D = 56$nm. By varying the contour length L_c of the chain our simulations indicate that the duration of unknotted events varies only modestly with L_c, while the average duration of knots (dominated by the relatively rare long-lived knots) grows significantly with L_c.

These properties can be rationalised by examining in detail the mechanism of knot formation and evolution along the chain. In particular it is found that chain unknotting depends mostly on the occurrence of significant back-folding of the chain ends, a process that is largely independent of L_c. On the other hand knots, although created/untied by the same back-foldings of the termini, subsequently move stochastically along the chain and hence their lifetime increases in parallel with the available contour length.

Based on these considerations one could predict that, at fixed D, the average knotting probability, which depends on the interplay of knots and unknots lifetimes, ought to grow with L_c mostly for the increased knots lifetime. This prediction is indeed in accord with the data of Figs. 3.6 and 3.7 and with the earlier conclusions of ref. [95] which addressed the dependence of the nanochannel width on the knotting unknotting dynamics of DNA strands of fixed contour length.

Chapter 4

Pore translocation of flexible chains with complex knots

In Chapter 3, I focussed on equilibrium knotting properties of DNA confined in nano-channels. In this chapter, instead, I will focus on an inherently out-of-equilibrium problem, that is the driven translocation through a nanopore of a knotted flexible polymer chain.

The motivation for this study is provided by ongoing efforts to use nanopore translocation as a means to sequence single-stranded DNA (ssDNA) filaments. The contour length of successfully-sequenced ssDNA runs has increased steadily in previous years [107] and pushing it significantly beyond the current limit of 50Kbps will inevitably lead to the frequent formation of knots. It therefore becomes of interest to map out how different knot types will impact the translocation compliance of model ssDNA chains.

This Chapter is organized as follows: after an introductory motivation and overview of the driven pore translocation of flexible chains, I will introduce our ssDNA model, which is based on a general model for flexible polyelectrolytes. I will then present the setup used to compare the translocation compliance of a wide range of knot types, included composite ones. I will finally show that the rich and seemingly disparate phenomenology can be captured in a seamless framework based on the mechanism by which the tractive force is propagated along and past the knots.

The content of this chapter is based on the work published in Ref. [58].

4.1 Introduction

The driven translocation of filamentous molecules through narrow pores is increasingly investigated in the context of polymer physics, molecular biology and single-molecule experiments [15, 72, 76, 107–126]. This translocation process naturally occurs in living organisms, where DNA [72, 127], RNA [128] and protein chains [129] are actively driven through biological pores, but is nowadays exploited to develop sequencing methods of single-stranded DNA filaments [107, 130–137] by measuring the ionic current blockaded [137, 138], or the electrical current perpendicular to the DNA backbone [135].

Because knots are inevitably formed in long equilibrated polymers [5, 45, 63, 139, 140], a relevant question is how the translocation compliance of a chain is affected by physical knots which are too cumbersome to pass through the pore [14, 15, 57, 67–69, 114, 141–145].

As a reference term we recall that several studies have shown that the time τ, required to translocate an *unknotted* chain through a narrow pore typically scales

as $\tau \propto N^{1+\nu}/f$ [109–111], where f is the dragging force, N is the number of chain monomers and ν is the metric exponent of a self-avoiding polymer [146].

How this behaviour is affected by the presence of knots, was first studied by Rosa *et al.* [142]. By considering the two simplest knot types, 3_1- and 4_1-knots, they found a surprising behaviour: while at high driving forces the presence of a knot jams the translocation process, at sufficiently low forces the translocation proceeds unhindered, and is indistinguishable from the case of unknotted chains.

The results of Rosa *et al.* posed several questions. For instance: beyond the simplest knot types, how does the translocation compliance vary with knot complexity? Also: what is the precise mechanism responsible for the onset of the topological friction, eventually leading to jamming, at high forces?

In this chapter, I will address these questions, especially regarding the mechanism by which the tractive force is propagated along and past the knots. The material presented is based on the study of Ref [58].

4.2 Methods

4.2.1 Model and simulation details

Inspired by the eletrophoretic driving of polyeletrolytes through nanopores [15, 107, 117, 121, 147–150] we model the translocating polymer as a flexible chain of beads with screened electrostatic self-repulsion. Specifically we considered chains of $N = 15,000$ identical beads of diameter σ and charge q.

In order to match single strand DNA (ssDNA) characteristics in a solution with 0.1M monovalent salt, it is possibile to set $\sigma = 1nm$, which is the nominal ssDNA thickness with each bead corresponding to about three nucleotides, and $q = -1.5e^-$ to the effective phosphates electrostatic charge, due to the counterions screening [151].

The chain internal energy consists of the sum of three terms: the chain connectivity term, U_{FENE}, the electrostatic screened self-repulsion, U_{DH}, and a truncated and shifted Lennard-Jones potential for the steric effects, U_{LJ}:

$$
\begin{aligned}
U_{\text{LJ}} &= \sum_{i,j>i}^{N} 4\epsilon_{i,j} \left[\left(\frac{\sigma}{d_{i,j}}\right)^{12} - \left(\frac{\sigma}{d_{i,j}}\right)^{6} + \frac{1}{4} \right] \theta(2^{1/6}\sigma - d_{i,j}) \\
U_{\text{FENE}} &= -\sum_{i}^{N-1} K R_0^2 \ln\left[1 - \left(\frac{d_{i,i+1}}{R_0}\right)^2 \right] \\
U_{\text{DH}} &= \sum_{i,j>i+1}^{N} \frac{q^2}{\epsilon_r d_{i,j}} e^{-\frac{d_{i,j}}{\lambda_{DH}}}.
\end{aligned}
\tag{4.1}
$$

$d_{i,j}$ is the distance of beads i and j, while ϵ is the energy unit and $\epsilon_{i,j}$ is equal to 2.33ϵ for consecutive beads ($j = i \pm 1$) and simply equal to ϵ otherwise. In the Lennard-Jones potential, θ is the Heaviside function, while in the FENE potential the stiffness is set to $K = 70\epsilon/\sigma^2$ and the maximum bond length is $R_0 = 1.5\sigma$. In the Debye-Hückel potential, ϵ_r is the dielectric constant and λ_{DH} is the Debye screening length.

The typical energy scale, ϵ, is set to match the system thermal energy $\epsilon = k_B T$, while $\frac{1}{\epsilon_r} = 1.6\frac{\epsilon\sigma}{q^2}$ and $\lambda_{DH} = \sigma$, respectively. For $\sigma \sim 1nm$, a typical coarse-graining level for polymers and biopolymers, these settings yield Bjerrum and Debye screening lengths of about 1nm, comparable to properties of water solutions with 0.1M monovalent salt at $T = 300K$.

All model parameters were set to match those of the earlier study of pore translocation of 3_1- and 4_1-knotted chain [142], except for the amplitudes of the bonding potential energy constants between consecutive beads. The latter have been reduced in order to decrease the statistical fluctuations of the bonding forces, which in equilibrium are roughly proportional to the square root of the potential curvature at the minimum. In particular the FENE elastic constant has been reduced from $K = 300\epsilon/\sigma^2$ to $70\epsilon/\sigma^2$, and the repulsive Lenard-Jones interaction between consecutive beads from 10ϵ to 2.33ϵ, so that the location of the bonding potential minimum remains unchanged. The limited bonding-force fluctuations are useful in the present context as they reduces the statistical error of the average traction forces that will be defined later on.
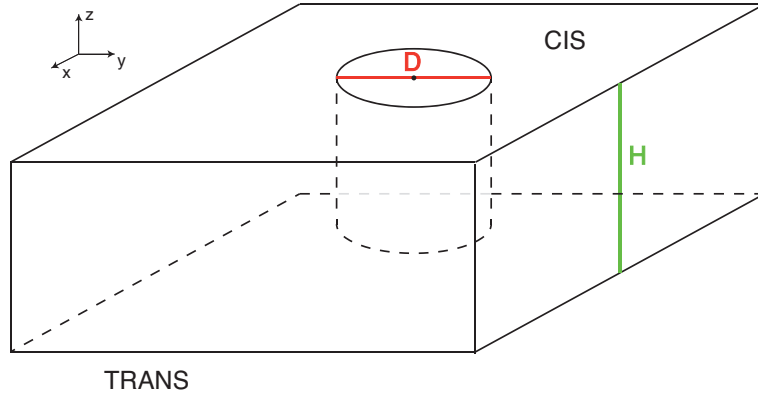

FIGURE 4.1: Schematic geometry of the nanopore of diameter $D = 2.8\sigma$ carved in a slab of height $H = 10\sigma$.

The pore is embedded in an impenetrable slab separating the *cis* (upper) and *trans* (lower) regions, see Fig. 4.1, and is narrow enough for only a single strand to pass through, thus preventing the passage of physical knots. The nanopore nominal diameter and longitudinal span are set equal to D=2.8σ and H=10σ, respectively, compatibly with experimental setups [151, 152]. The interactions with the slab and the pore are regulated by a truncated and shifted Lennard-Jones potential:

$$
\begin{aligned}
U_{pore} &= \sum_i^N 4\epsilon \left[\left(\frac{\sigma}{\delta_i} \right)^{12} - \left(\frac{\sigma}{\delta_i} \right)^6 + \frac{1}{4} \right] \theta(2^{1/6}\sigma - \delta_i) \\
U_{slab}^{cis} &= \sum_i^N 4\epsilon \left[\left(\frac{\sigma}{z_i} \right)^{12} - \left(\frac{\sigma}{z_i} \right)^6 + \frac{1}{4} \right] \theta(2^{1/6}\sigma - z_i) \quad \text{for} \quad x_i^2 + y_i^2 > \frac{D}{2} \quad (4.2) \\
U_{slab}^{trans} &= \sum_i^N 4\epsilon \left[\left(\frac{\sigma}{z_i + H} \right)^{12} - \left(\frac{\sigma}{z_i + H} \right)^6 + \frac{1}{4} \right] \theta(2^{1/6}\sigma - z_i) \quad \text{for} \quad x_i^2 + y_i^2 > \frac{D}{2}
\end{aligned}
$$

$$(4.3)$$

where $\delta_i \equiv \frac{D}{2} - \Delta_i$ and Δ_i is the distance of the ith bead from the pore axis. This setting leaves up a discontinuity in the borders of the pore, taken care by coating them with beads, kept fixed in their position, and interacting with the chain beads with their same Lennard-Jones potential.

The translocation is driven by a uniform longitudinal electric field active only within the pore and which pulls each bead inside it with a force f. The unit force $\frac{\epsilon}{\sigma}$

corresponds to about $4pN$ in real units.

The dynamical evolution of the chain is described with a constant-temperature Langevin dynamics, already mentioned in Chapter 3, and integrated with the LAMMPS [97] simulation package with default, standard parameters and with integration time step equal to $0.005\tau_{LJ}$, where $\tau_{LJ} = \sigma\sqrt{\frac{m}{\epsilon}}$ is the characteristic Lennard-Jones time unit. If we compare the diffusion of a single bead with Stokes law in water for a particle of $1nm$ diameter, τ_{LJ} corresponds to about $4.5ns$ in real unit. The Debye-Hückel potential has a cutoff set at 6σ.

The translocation is followed for a timespan sufficient to reach (and often overcome) the reference translocated fraction of the chain, $x = 0.3$. Depending on the force and knot topologies, the required timespan ranges from $10^3\tau_{LJ}$ to $10^7\tau_{LJ}$. For computing various observables we considered instantaneous snapshots sampled every $500\tau_{LJ}$.

All the results will be presented in simulation units.

4.2.2 Setup of initial configurations

The translocation process of a knotted chain, accommodating 3_1- or 4_1-knots, has been previously studied by Rosa *et al.* [142]. In equilibrated flexible chains, knots are located in the interior of the chain and can span thousands of monomers. For such typical configurations, at least two distinct regimes are observed as translocation progresses.

In the first regime, the mechanical tension propagates along the chain until it reaches the knotted region, and drags and tightens it at the pore entrance. Throughout this timespan, whose duration depends on the position and length of the knot, the translocation is essentially the same as for an unknotted chain.

After this first stage, the translocation can be hindered or jammed completely, if the dragging force causes a large enough friction in the tight knot trapped at the pore entrance (due to tightly interdigitating or interlocked chain monomers). The overall frictional effect would be qualitatively similar to the one observed in two tied ropes that are pulled apart [102].

These two distinct translocation stages are clearly observed in Fig. 4.2, which shows the translocation traces of two equilibrated chains tied in a 3_1- and 5_1-knot (generated with Monte Carlo sampling [27]).

For our study we wished to focus on the second stage, because it is dominated by topological friction, unlike the first stage that is more dependent on the chain geometry, which clearly varies from sample to sample. For this reason, we prearranged a tightened versions of the knots of our interest at the pore entrance, and attached to such leads (typically 100-monomer long) various instances of equilibrated – unknotted – chains generated with a Monte Carlo sampling procedure. A typical setup is illustrated in Fig. 4.3.

For each lead topology we considered at least 5 independent, equilibrated configurations of the attached chain. For the purpose of computing averages and statistical errors, data from different trajectories are treated as statistically independent. Finally, to avoid including the initial transient in the observables, the initial time interval of $5000\tau_{LJ}$ was disregarded.
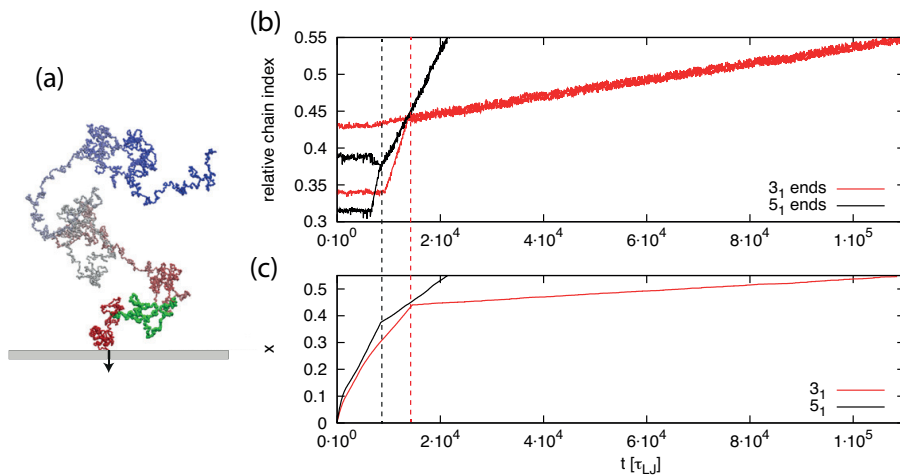
FIGURE 4.2: Driven translocation of equilibrated knotted chains accommodating 3_1- and 5_1-knots. The initial configurations of these knotted chains were obtained by Monte Carlo sampling. Hence, the position and length of the knots are dictated by canonical equilibrium. In particular, (a) shows the initial configuration for the trefoil knot, highlighted in green, before being pulled inside the pore through the slab, in gray. The translocation process at $f = 5\epsilon/\sigma$ is characterised by the time evolution of the translocated fraction of the chain (c), x, and of the position of the knot along the chain contour (b). The knot position is conveyed by the relative location on the chain contour of the two ends of the knotted region. For each knot type, two distinct translocation regimes are readily visible. The first one covers the time span required for the knotted region to reach the pore entrance, where it becomes tight. The duration of this timespan, where the behavior is the same as for unknotted chains, is different for the two chains because their knots are in different chain locations. After this transient, the translocation is hindered by the tightened knots. The different topological friction associated to 3_1 and 5_1 knots is readily noticed.
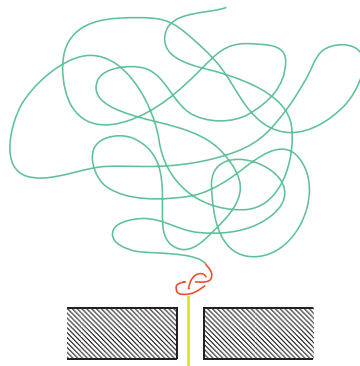

FIGURE 4.3: Initial configurational setup, composed in sequence by a short lead, up to 100-beads long (in yellow), a tight knot (in red), and an equilibrated unknotted chain (in green).

4.2.3   Considered knot types

The considered repertoire of knots is sketched in Fig 4.4 and extends significantly beyond the 3_1 and 4_1 instances considered in Ref. [142]. The set is ordered for increasing nominal complexity, i.e. the number of crossings in the minimal two-dimensional projection, which is generally a key determinant of topological friction [100–105, 114, 153, 154].
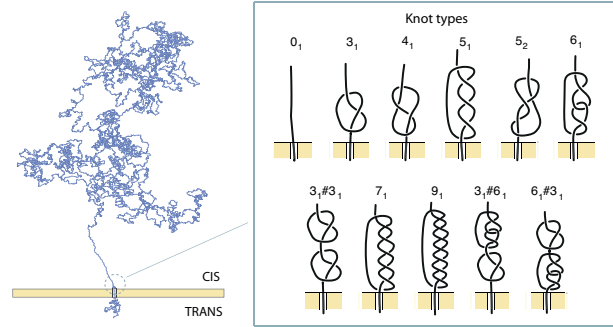

FIGURE 4.4: Snapshot of an early stage of translocation of a knotted
chain. Inset: repertoire of knots prepositioned at the pore entrance.

The repertoire of knots includes a sizeable number of representatives from the two main topological families of simple prime knots: twist knots and torus knots, introduced in Chapter 1. The former group includes the 3_1, 4_1, 5_2 and 6_1 knots, while the latter includes the 3_1, 5_1, 7_1 and 9_1 knots. Notice that the 3_1 topology appears in both classes.

In addition, we consider composite knots resulting from the serial addition of prime knots. We recall that composite knots are the most common form of entanglement in sufficiently long chains [27].

The position of knots was detected in the usual way using the Alexander polynomial and the bottom-up approach to identify the shortest portion accommodating the chain.

4.3   Results and discussion

4.3.1   Translocation compliance for different knot types

The translocation compliance across the knots repertoire was studied for values of f in the 0.1–8 ϵ/σ range.

The results are shown in Fig. 4.5 and pertain to the average translocation velocity, v, measured at $x = 0.1, 0.3, 0.5$, with x being the chain fraction already passed through the pore. The behaviour is analogous independently of the reference chain fraction.

The reference case of unknotted chains is always shown in a separate panel, because of its wider range of velocities, and v has the expected linear dependence on f at fixed x [109].

By contrast, the knotted chain translocation speed presents several unexpected tiers of complexity upon varying both the applied force and the knot type.

First, at any of the considered forces, the average velocities do not systematically decrease with knot complexity. Secondly, only few of the prime knots display the non-monotonic v versus f dependence signalling the onset of jamming. These are
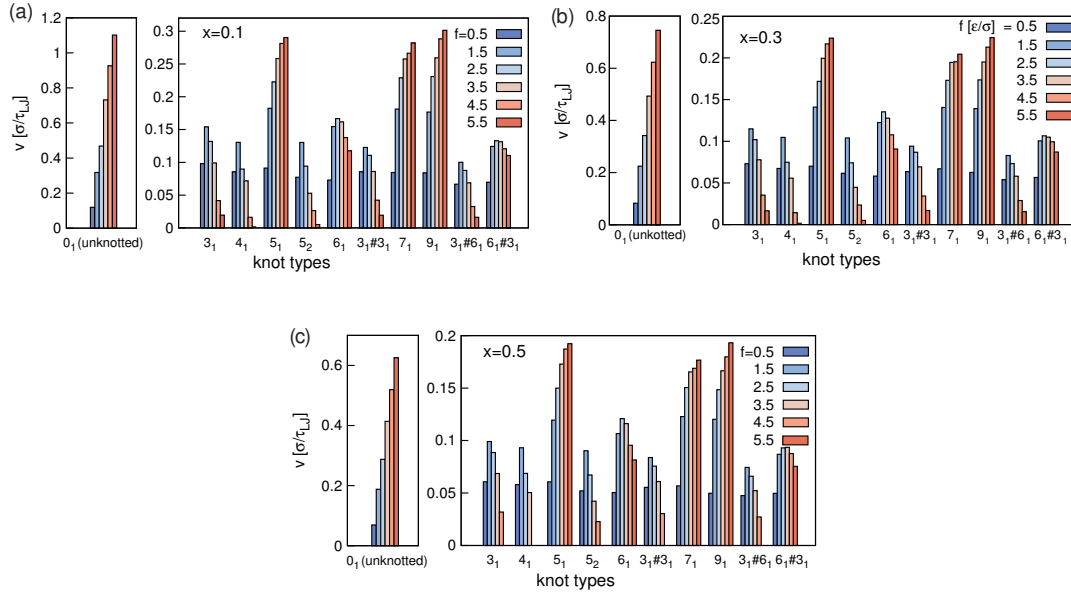
FIGURE 4.5: Average translocation speed, v, measured, at 10%,30% and 50% translocated fraction, for various applied forces, f. Across all cases the relative error on v is always smaller than 10% and equal to $\sim 5\%$ on average. The behaviour is analogous in the three cases shown. Only minor differences are discernible for the $x = 0.1$ case when the translocation time is short enough to compete with the initial transient during which the system adjusts to the sudden application of the electrophoretic force at the beginning of the simulation. Also, because of the exceedingly slow translocation process of some jammed knots, the data for some forces are not present for $x = 0.5$ due to incomplete translocations.

knots belonging to the twist family. Even increasing the force applied to each of the about 10 beads in the channel up to $f \sim 8\epsilon/\sigma$, the largest that FENE bonds can withstand in the simulation protocol, the onset of jamming is not seen for the 5_1, 7_1 and 9_1 torus knots. To our knowledge, this interesting qualitative dicotomy has been previously found only in a specific context, namely for low- or moderately-tensioned knotted chains in a surrounding fluid (with 3_1 knot behaving as a torus rather than as a twist knot) [153, 154]. In these cases torus and twist knots still have comparable mobilities, while here their high-force translocation velocities can vary by an order of magnitude, see Fig. 4.5.

Finally, the translocation behaviour of compositely-knotted chains is unexpectedly rich and intriguing, too. In fact, their hindrance can be comparable and even lower than for chains hosting only one of their components, see e.g. $3_1\#3_1$ vs 3_1 and $6_1\#3_1$ vs 3_1, and it is distinctly non-commutative in the knots order, see $6_1\#3_1$ vs $3_1\#6_1$. Morover, the sequential order of the initially separate and tigthened prime components is preserved at all times in our simulations. In fact, the applied forces and pore/slab impenetrability prevent the prime components from fluctuating in size enough to pass through each other, unlike what can happen in untensioned chains [155].

We notice that the considered forces are all large enough to have prime knots stay closely localised at the pore entrance. This is clearly illustrated in Fig. 4.6 which shows the distance of the first bead of the knotted region from the pore (Fig. 4.6a)

and the contour length of the knotted region too (Fig. 4.6b). For composite knots, the first knot component nearest to the pore remains tightened, while the second one feels a lower pulling force increasing f. This is reflected by observing a growth in the knot length, see Fig. 4.6c.
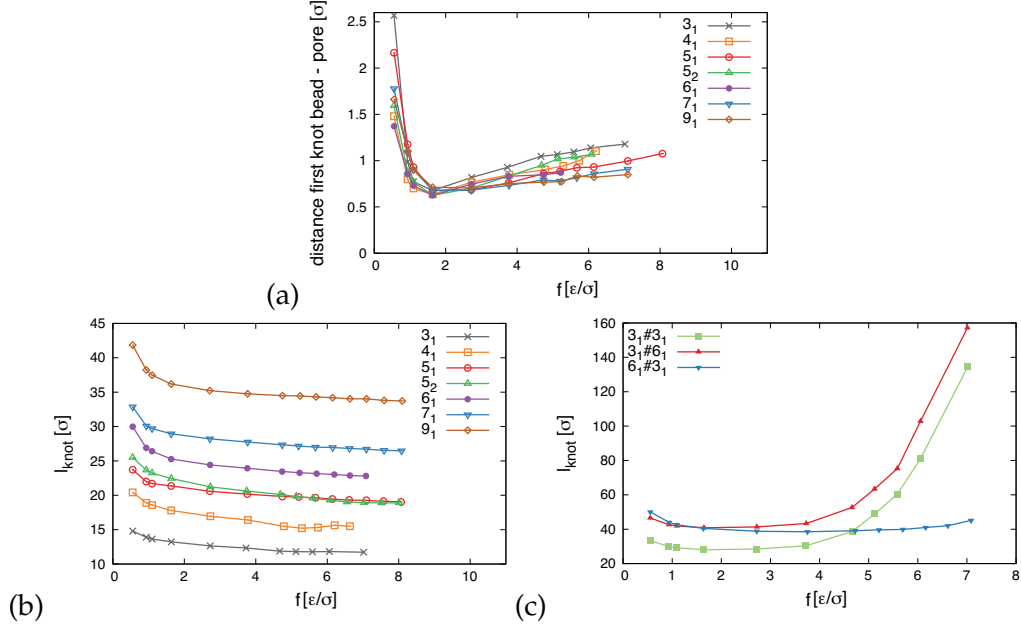

FIGURE 4.6: (a) Distance from the pore of the first bead of the knotted region and (b) knot length for prime knots as a function of the pulling force f. The contour length of few *composite* knots is shown in panel (c).

4.3.2 Traction force before and past the knot

These seemingly disparate properties can be quantitatively accounted for by a seamless framework based on how the driving force is propagated along and past the knot trapped at the pore entrance.

Following the spirit of the Euler-Eytelwein analysis of tension and friction in a rope wound around a capstan [156] we profile the relationship between the chain tractive force measured immediately before and after the knot (or simply before and after the pore for unknotted chains). These two key quantities will be denoted with t_{in} and t_{out} respectively. The former is obtained by multiplying f by the average number, $n \sim 10$, of beads in the pore: $t_{in} = n f$, and is clearly constant during the translocation process. t_{out} is instead the average tractive force experienced by chain beads immediately past the knot and is measured directly. For one of these beads, with index i, the instantaneous traction is obtained by computing the total deterministic force exerted on it by all preceding beads, $1, 2, \dots i - 1$ (index 1 is for the chain terminal primed in the pore lumen at the beginning of the simulation) and by projecting it on the $i, i + 1$ bond direction.

To reduce statistical errors, t_{out} is averaged over the simulation time and over the first ten beads of the region immediately past the knot (or past the pore for unknotted chains). The error on t_{out} is estimated by treating as independent the average tractive forces computed for each run. The fact that t_{out} does not vary appreciably during the translocation is confirmed *a posteriori*, see Fig. 4.7.
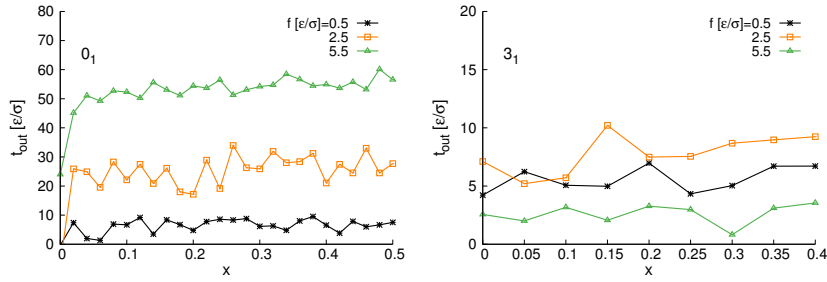
FIGURE 4.7: t_{out} measured at fixed translocated fraction, x, for un-knotted and trefoil-knotted chains.
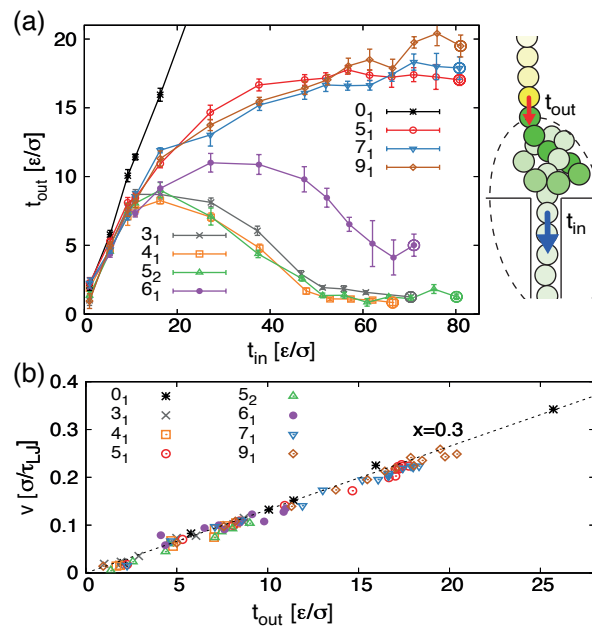

FIGURE 4.8: (a) Average chain tractive force at the beginning (t_{in}) and end (t_{out}) of various prime knots. Beyond the circled data points chain breaks occur systematically. Sketch: the tractive force, t_{out}, on the yellow bead immediately past the knot is given by the total longitudinal force exerted on it by the preceding part of the chain (circled, green). (b) Translocation speed at 30% translocated fraction versus t_{out} for all considered prime knots.

The relationship between t_{in} and t_{out} is shown in Fig. 4.8a for all considered prime knots. The significant reduction of the transmitted traction operated by any of the prime knots vividly emerges from the comparison with the unknotted case. More important, however, is the substantial spread of the output traction across the knots repertoire. This heterogeneity originates from the very diverse response of torus knots from twist ones. It is clearly the latter that produce the largest topological friction and, due to the rapid decrease of t_{out} with t_{in}, can eventually halt the translocation process at the largest applied forces, unlike torus knots.

The force transmission, or force reduction, curves, which to our knowledge have not been previously introduced in this context, are key to characterize the varied

behaviour of prime knots in Fig. 4.8a, and to predict that of composite ones.

To this end we note that the complex translocation phenomenology across all knots in Fig. 4.8a is underpinned by a simple linear relationship between v and t_{out}, see Fig. 4.8b. Notice that the data include the case of unknotted chains too. Therefore, the behaviour of chains accommodating very different knots is practically equivalent to that of an entanglement-free chain translocated by a total force equal to t_{out}.

4.3.3   Tension and strain inside the knotted region

The detailed mechanisms governing the dissipative transmission of traction along the knot contour can be gleaned from the bond strain profiles of the chain portion near the pore.

To compute the bond strain profile we first consider each instantaneous configuration and assign to each bead a sequence index *relative* to its pore proximity. Specifically, for each sampled configuration, the *cis* bead closest to the pore entrance is assigned index 0; the other beads are then sequentially indexed, with *trans* bead having negative indexes, and *cis* beads having positive ones. Clearly, this relative indexing is dynamical, in that it changes during the translocation process. The bond strain profile is then computed by averaging, over all sampled configurations, the length of bonds connecting beads with the same relative indexing.
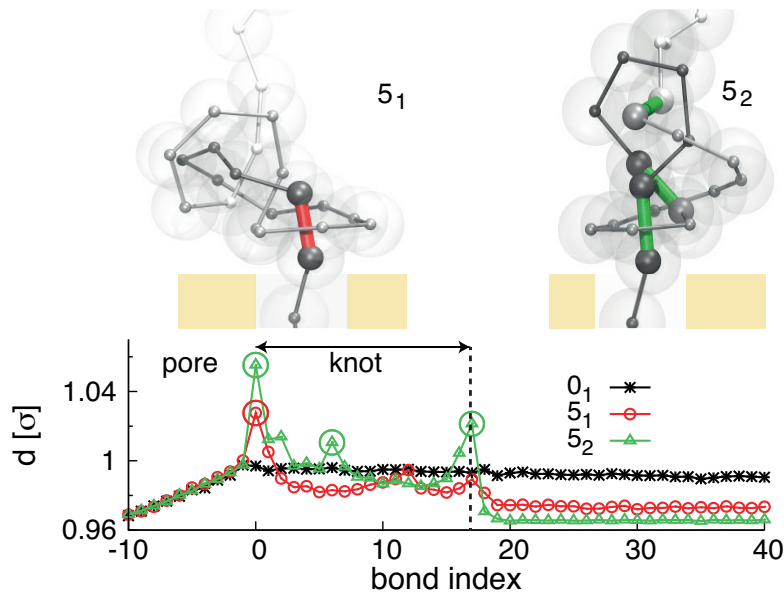

FIGURE 4.9: Profile of the average length, d, of chain bonds proximal to the pore for three types of knotted chains pulled at $f = 5.5$; index zero denotes the first *cis* bond. Bonds with the peak elongation (circled) are highlighted in the snapshots.

Typical examples are shown in Fig. 4.9 and additional ones are provided in Fig. 4.10. Compared to the smoothly decreasing strain of the unknotted case, the strain profile is riddled with peaks in the knotted region and falls to negligible values immediately outside it. Unlike equilibrated chains pulled at both termini [157], the peaks profile is noticeably asymmetric with respect to the knot midpoint, with the highest strain occuring almost invariably at the pore entrance.

Considering that the bond strain reflects the chain self-friction along the knot contour, one may envisage that the largest dissipation of the input traction should occur at the point with the highest strain.
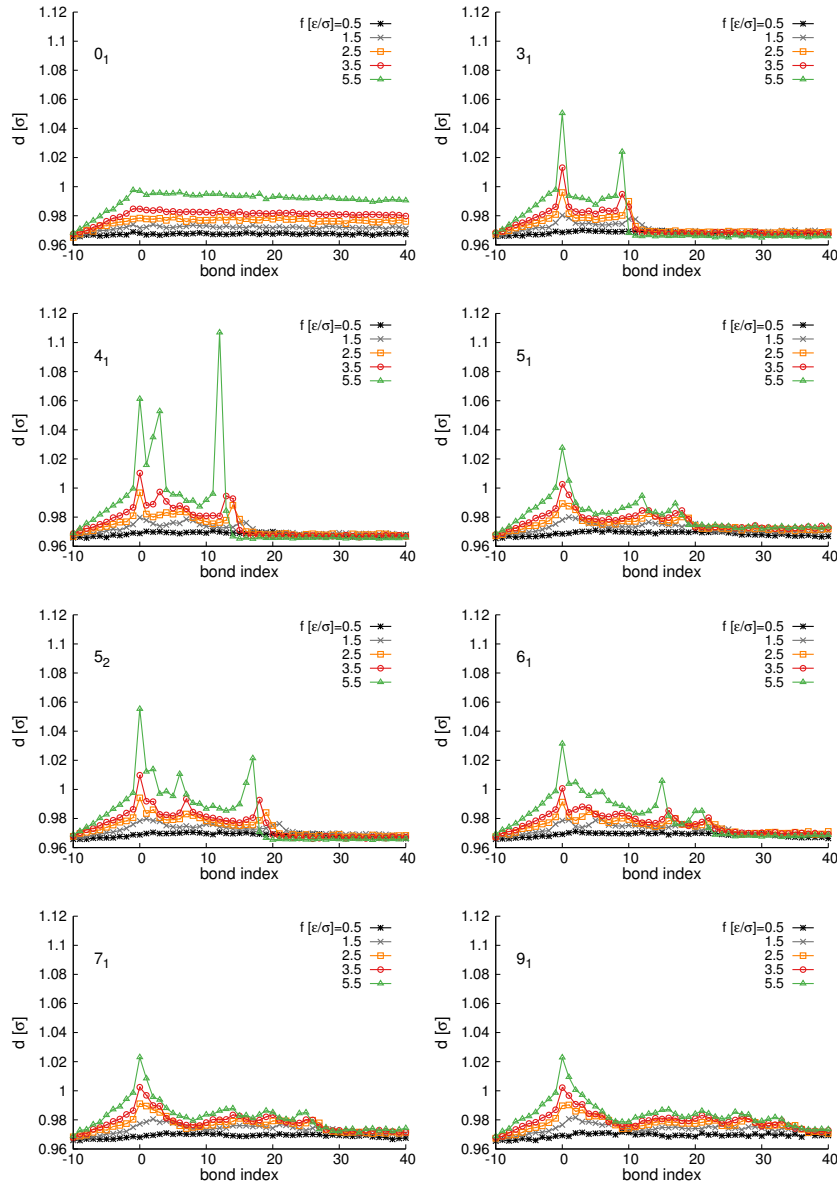

FIGURE 4.10: Strain profiles for different prime knots and for unknotted chains too. The same scale is used in all plots.

This is confirmed by Fig. 4.11a where the topological friction, as captured by the traction reduction coefficient, $r \equiv t_{out}/t_{in}$, is plotted against the maximum (peak) bond extension, d_{max}. Again, the behaviour of torus and twist knots are neatly separated, and the dispersion within either of the two families is much reduced compared to Fig. 4.8a.

Mapping out the strain peaks on the actual conformations of the tightened knots clarifies the origin of the major difference between torus and twist knots. As exemplified by the snapshots shown in Fig. 4.9, torus knots have a single region, namely
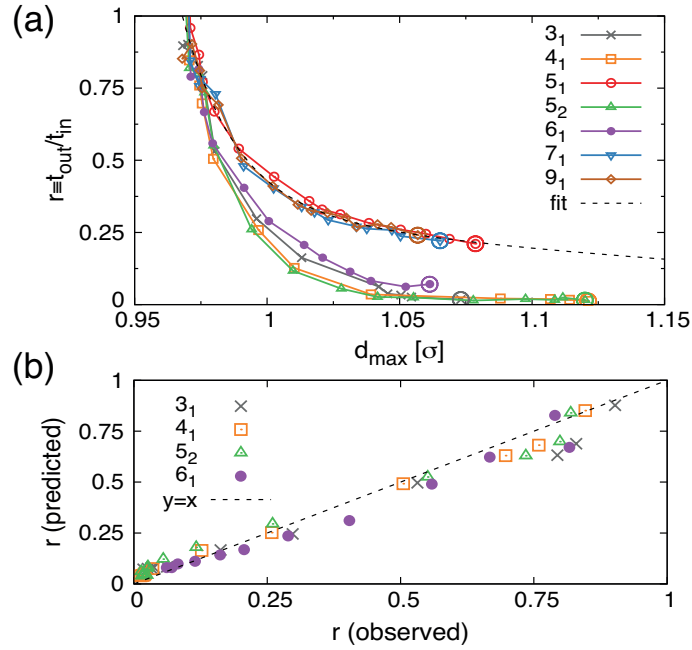
FIGURE 4.11: (a) Traction reduction ratio, $r \equiv t_{out}/t_{in}$, versus the maximum bond length, d_{max}. The dashed line is the fit of the torus knots data points using, empirically, a double-exponential: $r(d) = 0.57 \exp[-(d - d^0)/0.020] + 0.34 \exp[-(d - d^0)/0.23]$ with $d^0 = 0.97$. (b) Predicted versus measured traction reduction ratio for twist knots.

the knot entrance, where the strain peaks overcome the reference profile of the un-knotted case. Twist knots instead have two distinct regions: the knot entrance plus the exit.

This suggests that the much higher hindrance offered by twist knots results from the combination of two distinct dissipative mechanisms, as opposed to the single one of torus knots.

To verify this hypothesis we parametrize the functional dependence of the traction reduction ratio of torus knots, r versus d_{max}, see dashed line in Fig. 4.11a, and used it to predict the traction drop of twist knots. Specifically, we assumed for simplicity that the two distinct friction regions are independent and acting as the single friction point of torus knots. Accordingly, the predicted traction reduction of twist knots is given by the multiplicative composition of the individual frictional effects: $r_{pred} \equiv r(d_1) \cdot r(d_2)$ where d_1 and d_2 are the peak bond elongations at the knot entrance and exit.

As shown by Fig. 4.11b, the predicted traction reduction is in very good agreement with the measured one. We note that the predictions are practically deterministic, since no tunable parameter is involved. This further supports the use of the traction reduction at highly-strained points as the dominant mechanism for capturing and unifying the observed disparate phenomenology of Fig. 4.5.

Indeed, the same concept can be used to predict, and hence rationalize the two key features of the overall hindrance offered by composite knots, namely that it can be smaller than that of their individual components and that it is non-commutative, see Fig. 4.5.

Notably, both aspects are straightforwardly recovered from the peak strain analysis or equivalently and more simply, from the traction transmission curves of prime
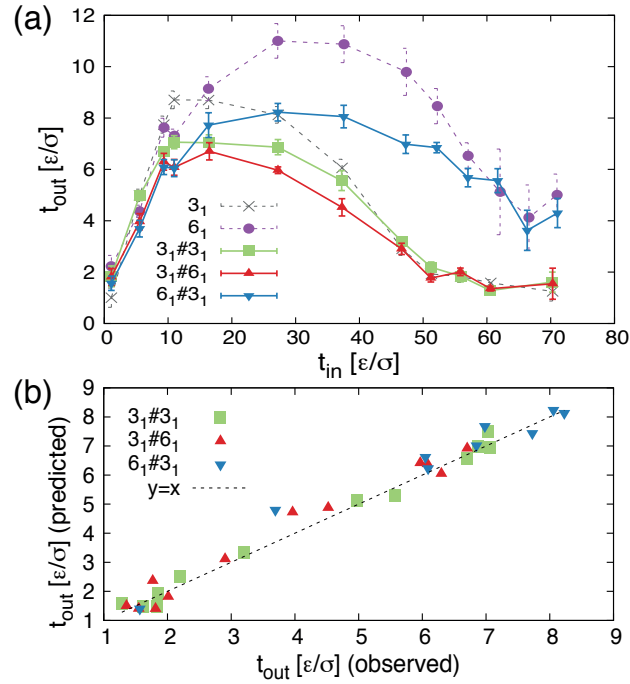
FIGURE 4.12: (a) Average chain tractive force at the beginning, t_{in}, and end, t_{out}, of composites knots and their components. (b) Predicted versus measured t_{out} for these knots.

knots.

4.3.4 Prediction for composite knots

The same concepts used for the predictions for Fig. 4.11 can be transferred to composite knots, by assuming that the global output traction of composite knots, shown in Fig. 4.12a, mostly results from the multiplicative combination of the frictional effects of their prime components when these are tightened separately and kept in contact near the pore entrance by a pulling force.

For instance, for given f (and hence t_{in}) the predicted output traction of the $3_1\#6_1$ knot is obtained by using the curves in Fig. 4.8a to (*i*) establish the tractive force $t_{out,3_1}$ immediately past the first (3_1) knotted component, then (*ii*) using it as t_{in} for the second (6_1) component, and finally computing the corresponding t_{out}, which yields the predicted output force. Notice that this scheme is intrinsically non-commutative and hence has *a priori* the potential of accounting for both of the aforementioned effects. Indeed, the predicted output forces are in very good agreement with the actual measurements for all considered composite knots, see Fig. 4.12b.

We emphasize that the predictions are based only on a spline interpolation of the data in Fig. 4.8a with no free adjustable parameters.

4.3.5 Conclusions

In conclusion, we have shown that the complex and varied phenomenology of knotted chains translocation can be seamlessly captured in a transparent, quantitative framework based on the mechanism by which the tractive force is propagated along

and past knots. This scheme, here studied for a generic model of flexible polyelectrolytes, ought to be useful in applicative contexts such as nanopore translocation of nucleic acids filaments [15, 107, 117, 121, 137, 147–150], especially the flexible single-stranded ones. In particular, it could be useful for inferring the general topology of chains based on their translocation response or for harnessing their topological friction to optimally reduce and control their translocation speed.

Chapter 5

Pore translocation of knotted DNA rings

In this Chapter, I will dwell further on the driven translocation of knotted polymers through a nanopore, but I will shift the focus from fully-flexible chains (Chapter 4) to semi-flexible ones.

As I discuss below, the main motivation for such extension is a recent set of groundbreaking experiments that have allowed Dekker and co-workers to profile with unprecedented detail the properties of knotted DNA molecules pulled through solid-state nanopores [15]. More precisely the DNA knotting features were established indirectly from the time-modulation of the ionic current flowing through the pore.

Here, I will report on the use of a detailed mesoscopic DNA model to reproduce *in silico* the typical experimental conditions and show how the results can be used to improve substantially the interpretation of experimental measurements as well as indicating promising directions for more advanced characterizations.

The results I will present have been recently published in Ref. [158], which provides the basis for the present chapter.

5.1   Introduction

In a breakthrough experiment [15], Plesa *et al.* used an electric field to drive spontaneously knotted DNA chains of 2, 20, 50 and 160 Kbps through solid state nanopores with diameter between 10 and 20nm, measuring the recurrence of knots. Unlike previous techniques, such as gel electrophoresis, which can be used to identify knots in DNA chains of up to 10Kbps, nanopores can in principle be used to detect knots in DNA filaments of arbitrary length.

In the experiments, the pore size was purposely chosen to be smaller than the DNA persistence length, $l_p = 50$nm, and yet wide enough to accommodate several double-stranded DNA (dsDNA) strands, so that knots could pass through it rather than being trapped at the pore entrance.

A surprisingly rich phenomenology emerged from the analysis of the main monitored observable, namely the time-dependence of the ionic current amplitude. As it is shown in Fig. 5.1, the inspection of the ionic current modulation allowed for timing when the chain entered and exited the pore, whether the chain translocation was accompanied by a knot-passage event, when exactly the latter occurred and how long it lasted.

These measurements carry a wealth of information about the knotting properties of DNA. Arguably, the most direct information regards the DNA knotting probability in equilibrium, which is given by the percentage of chain translocation events that
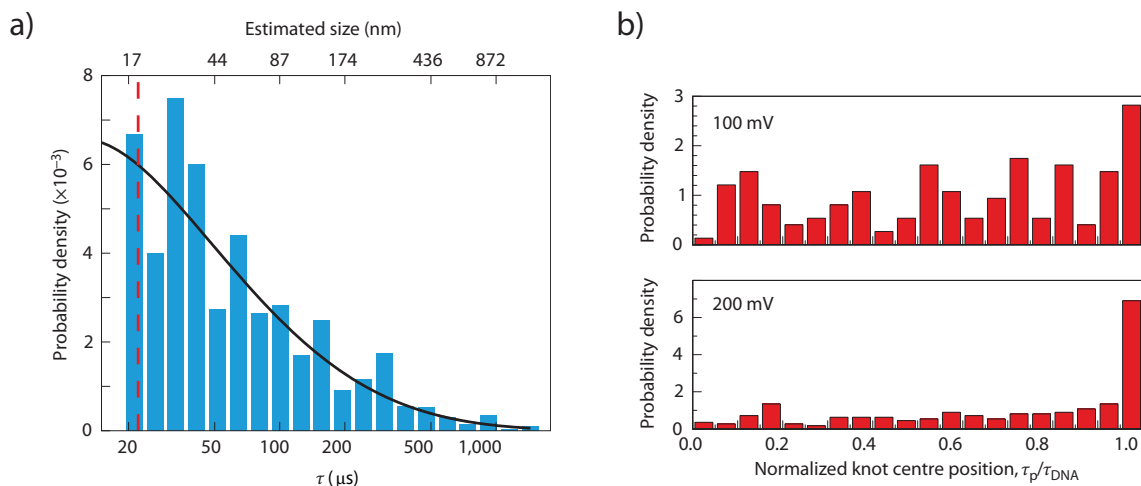
FIGURE 5.1: a) Probability distribution of the elapsed time at which the pore was obstructed by the passing knot. b) Duration of the obstruction event. From [15].

bear the distinctive drop in ionic current that is the hallmark of a (multistrand) knot passage event. Other properties of interest, and particularly the contour length of the knotted region and its positioning along the chain – for which theoretical predictions have been made [159] – can be related only indirectly to the timing measurements.

These considerations motivated our study [158], based on molecular dynamics simulations of oxDNA, an accurate mesoscopic DNA model [160–164]. Specifically, we investigated in detail the translocation process and put forward novel interpretative schemes for the associated interplay of DNA geometry and topology.

5.2   Methods

5.2.1   Model

For an accurate, mesoscopic description of double-stranded DNA we used oxDNA [161–164]. In this model, depicted in Fig. 5.2, each nucleotide is described by three interaction centers kept in a precise, rigid, geometrical relationship. The potential energy includes terms that account for the chain connectivity, bending rigidity, base-pairing,
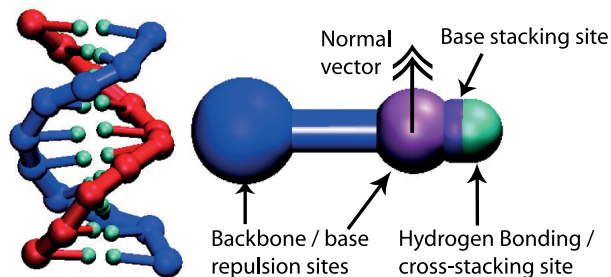

FIGURE 5.2: On the left, a duplex as represented by the oxDNA model. On the right, detailed structure of a single nucleotide, composed by three interaction centers rigidly connected. From [160].

screened electrostatic interactions and stacking effects. These terms are parametrised to reproduce the salient structural and equilibrium properties of nucleic acids filaments, such as the helical pitch, persistence length, and torsional stiffness of double-stranded molecules, at various values of the system temperature and salt concentration, here set to $T = 300$K and 1M NaCl, respectively.

The simulation unit length corresponds here to $\sigma = 0.85$nm. The nanopore is embedded in a slab 10nm wide and 10nm long, parametrized with the same potential used in Chapter 4.

We carried out Langevin dynamics simulations at constant-temperature $T = 300$K, without hydrodynamic effects, with a time step of $0.005\tau_{LJ}$, where τ_{LJ} is the standard Lennard-Jones time unit for the simulations. We set the simulation parameter as the system's default.

An approximate mapping with real time can be obtained by measuring the diffusion coefficient of a small oxDNA fragment of 4bps $D = 0.31\frac{\sigma^2}{\tau_{LJ}}$, with shape resembling a sphere of about 1.27nm, and matching it with the diffusion coefficient in water for an actual bead of the same diameter $D_{stokes} = 3 \cdot 10^{-10}\frac{m^2}{s}$. The comparison yields $\tau_{LJ} \sim 0.7$ns.

The translocation is driven by a longitudinal electric field exerting a force of 0.2pN on each nucleotide inside the pore. This value was selected after the experiments of Ref. [165], where the total pulling force applied on a 48Kbps chain through a pore, with a 100mV electric field and in 1M NaCl solution, was about 24pN. Because the average number of nucleotides occupying the pore at any given time is 120, this correspond to have each nucleotide pulled with 0.2pN.

For simplicity, we neglect the action of the field outside the pore [166, 167] that, in actual realizations, can facilitate the capture and pore insertion of the knotted chains [112, 168, 169].

5.2.2 Setup of initial configurations

The initial conformations of the 10Kbp-long DNA rings to be translocated through the pore were obtained with a three-tier modelling procedure.

First, we used a topology-unrestricted Monte Carlo scheme to sample efficiently the conformational space of 10kbp-long DNA rings in equilibrium in bulk and then extract *a posteriori* tens of distinct (uncorrelated) knotted instances.

For such stochastic exploration we resorted, as is customary [12, 27, 170], to modelling the DNA rings as rings of 340 cylinders. The length of the cylinder axis was set equal to 10nm, while the cylinder diameter was set to 2.5nm, see Fig. 5.3a. Non-consecutive cylinders are subject to excluded volume interactions, while consecutive cylinders interact via a bending energy term whose strength is appropriate to reproduce the know DNA persistence length, $l_p = 50$nm. At these ring lengths, the knot spectrum is vastly dominated by trefoil knots [62, 145, 170], and hence we singled out from the MC sample tens of uncorrelated trefoil-knotted rings.

These rings of cylinders were then subjected to a first level of fine-graining, by lining the centerline of the cylinders with contacting beads of nominal diameter equal to 0.34nm, the typical basepair distance. The resulting ring of beads, whose potential energy again included steric interactions and a bending rigidity matching the DNA known persistence length, was next relaxed with a Langevin molecular dynamics for a time interval of ($10^4\tau_{LJ}$). The stochastic dynamics was integrated with the LAMMPS software package, with a protocol analogous to that described in Chapter 3. An example of the resulting relaxed configurations is shown in Fig. 5.3b.
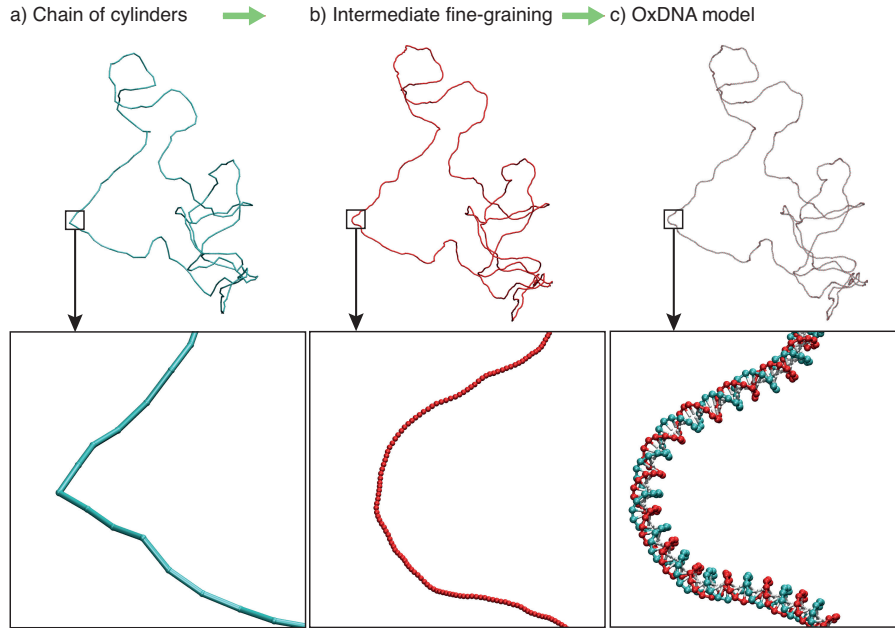
FIGURE 5.3: Three-step procedure to obtain the initial conformations.
a) Chain of cylinders, obtained by the Monte Carlo sampling, with
each cylinder corresponding to about 30 nucloetides. b) Chain of
beads, with each bead corresponding to a single base-pair. c) OxDNA
starting configuration, formed by a double-helix.

The ring of beads were finally fine-grained at the level of the oxDNA model, by interpolating the centerline of the chain of beads with two model DNA strands, see Fig. 5.3c.

Note that, because the experiments were carried out on nicked DNA, the model oxDNA configurations that we sought to generate had to be torsionally-relaxed. For unknotted rings arranged in a planar circle, this prescription amounts to introducing 1 turn of the double helix every 10.5bp. For our 10kbp-long rings, this corresponds to a total twist of $Tw = 952$ turns.

For a general unknotted chain, placed on a plane, Tw also coincides with the total linking number L_k, because the writhe amounts to zero $Wr = 0$ and the general formula $L_k = Tw + Wr$ holds [171]. In the general case of unknotted non-planar configurations, however, the writhe will be different from zero, and the twist needs to be computed as $Tw = L_k - Wr$.

For knotted chains, that can never be arranged as a planar circle, a further correction term to the formula connecting link, twist and writhe is needed [172], and corresponds to the subtraction of the chain's average writhe [173].

We therefore used these prescriptions to impose the correct number of turns on each chain of beads, to obtain torsionally-relaxed double-stranded oxDNA rings. The writhe was calculated using the numerical recipe of Ref. [174]. The global twist was equally divided along all the bases, with the helix constructed again using the method of Ref. [174].

To mimic experiments, such rings were finally nicked by removing one base from a single base-pair of the chain, as in typical experiments [15, 165]. This implies that only a ssDNA filament has a full chain connectivity. We verified *a posteriori* that the oxDNA rings remained structurally stable during the simulation, and did not fray at the nicks.
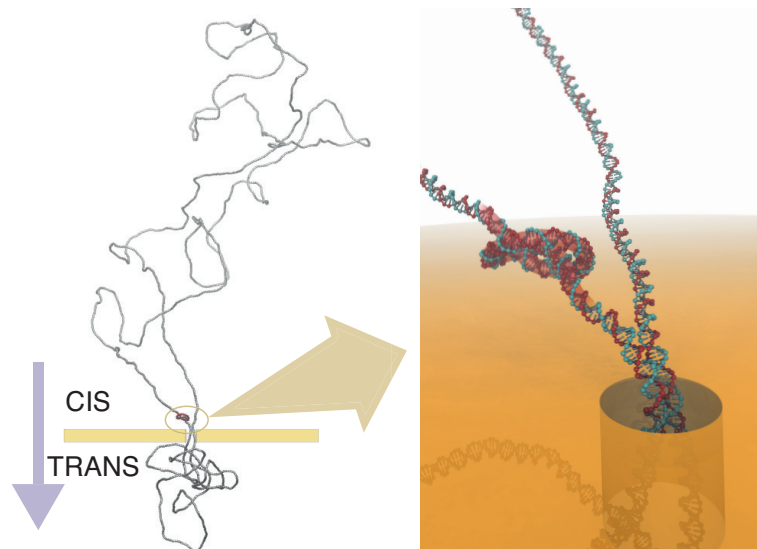
FIGURE 5.4: Typical configuration of a knotted double-stranded DNA ring translocating through a nanopore. The knot approaching the pore entrance is highlighted in the inset. The ring, which is modelled with oxDNA, is 10Kbp-long and the pore, which is 10nm wide, is embedded in an impenetrable 10nm-thick slab. A translocating force of 0.2pN is applied to each nucleotide inside the pore.

The lowest z-axis nucleotide was chosen, for simplicity, as the starting root of the translocation process, and was positioned over the pore by translating the whole chain. We then applied a small extra electric field outside the pore, in order to pull the first base-pairs inside it, and turned it off during the normal translocation procedure.

The translocation dynamics was studied for 50 different equilibrated knotted DNA configurations. A typical configuration is shown in Fig. 5.4. The simulations were carried out using GPUs [163], and each run required about 100 computational hours.

5.2.3 Observables

The passage of a knot, or of its essential crossings, through the pore was revealed by monitoring the number of nucleotides inside the pore and detecting increases by more than 30% from the baseline value, which is about equal to 60bp. This threshold criteria, which was validated by supervised visual inspection, was also used to establish the duration of the time interval associated to the passage of the knot through the pore.

For each instantaneous configuration, the location of the knot was identified with a bottom-up search, described in Chapter 1, identifying the shortest portion of the ring that, after suitable closure, has the same topology of the original ring. The search is limited to the *trans* or *cis* parts of the rings, respectively, depending on whether the knot has or has not already translocated through the pore.

5.3   Results and discussion

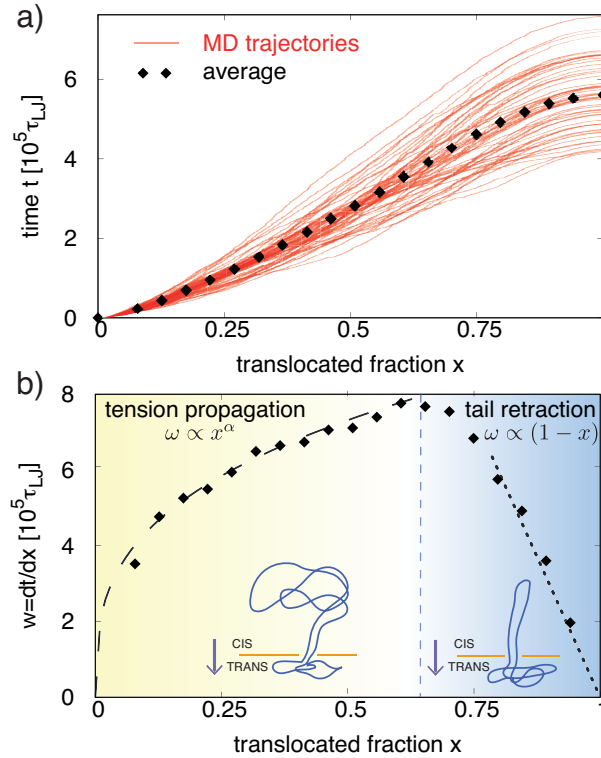### 5.3.1   Translocation dynamics overview

FIGURE 5.5: a) The time required to translocate a fraction x of the knotted DNA rings is shown for 50 independent simulations (red curves). The black points show the average $\langle t \rangle$ versus x curve. b) The waiting time curve, $w = \frac{d\langle t \rangle}{dx}$ highlights two main regimes corresponding to the tension propagation along the chain ($x \lesssim 0.5$) followed by the translocation of the rectified chain tail ($x \gtrsim 0.75$). The dashed and dotted lines are best fits based, respectively, on $w \propto x^\alpha$ (yielding $\alpha \sim 0.32$) and $w \propto (1 - x)$.

For a first, general characterization of the process we profiled the translocated fraction of the chain, x, as a function of the elapsed simulation time, t. This dependence is shown in Fig.5.5a where, as customary, it is presented as a t versus x plot. The red curves cover the individual trajectories while the filled points represent the average curve.

Fig.5.5b shows, instead, the so-called waiting time [175], $w = dt/dx$, i.e. the inverse of the translocation velocity, whose profile clearly outlines the two known main translocation regimes [112, 175, 176].

The first part of the curve, for $x \lesssim 0.5$, corresponds to the tension propagation along the chain, which itself presents an articulate phenomenology [177–179]. For chains that are asymptotically long and free of entanglement, theoretical scaling arguments predict a power-law behaviour, $w \propto x^\nu$, where $\nu = 0.586$ is the metric exponent for self-avoiding walks [109–111, 175, 177, 180], while smaller effective exponents are expected for chains of finite length [175, 181]. Fig. 5.5b shows that data points for the 10Kbp chains are indeed well fitted by a power law, and the

effective exponent, 0.32, is close to what previously reported at comparable DNA lengths [182].

At $x \sim 0.6$ the tension propagation regime crosses over to the tail retraction regime. In this stage the still untranslocated remainder of the chain, which is fully rectified, accelerates towards the pore. Because the pore is large enough to let the whole knot pass through, this second stage too follows the behaviour expected theoretically, $w \propto (1 - x)$ [175].

In real time units, the typical translocation time for the 10Kbps DNA chain is $t = 400\mu$s. Using the same scaling law $t \propto x^{1+\nu}$, we can extrapolate this time to 5ms for longer chains of 50Kbp, which compares well with experimental measurements available for this chain length [165].

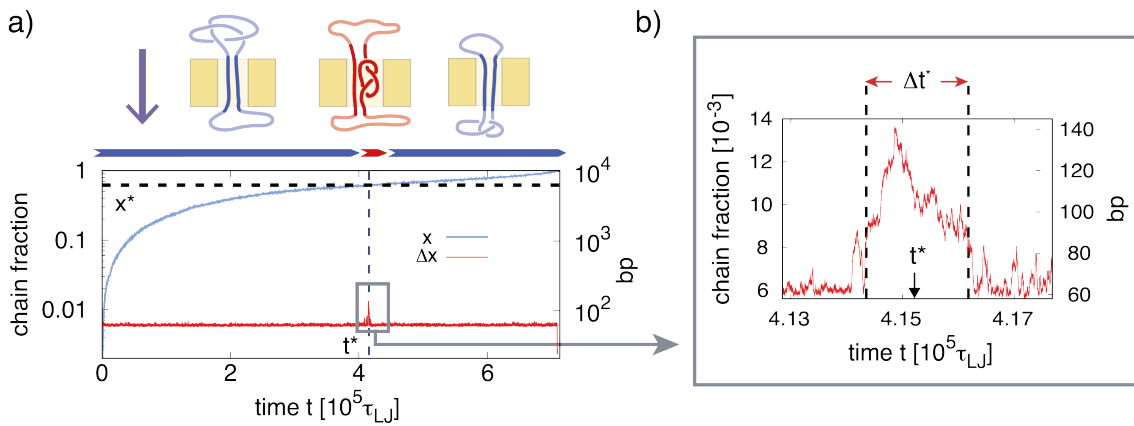### 5.3.2 Statistics of knot translocation events

FIGURE 5.6: a) Time evolution of the chain fraction that is inside the pore, Δx (red), and that has already translocated through it, x (blue). An absolute scale in base-pairs for x and Δx is also provided for the semi-log plot. The knot passage event is highlighted in panel b) and its time of occurrence, t^*, is defined as the midpoint of the time interval Δt^* during which Δx exceeds by more than 30% its baseline value.

Inspired by experiments, we detect the passage of the knot through the pore by monitoring the degree of obstruction of the latter. During such event, in fact, the pore lumen must accommodate up to four double-stranded filaments, instead of the usual two, see sketches in Fig. 5.6a.

We accordingly monitored the time evolution of the chain fraction inside the pore, Δx, which is shown in Fig. 5.6a. The pore obstruction caused by the passing knot is indeed signalled by a bump that stands out against the Δx baseline, see Fig. 5.6b. Notice that this major pore obstruction event is preceded by a smaller signal burst caused by the knot partially entering the pore and then retracting from it. Such translocation attempts affect about 50% of the trajectories. Their occurrence arguably depends on frictional effects arising from the geometry of the knot and the direction with which it engages the pore.

Various observables of interest, related to those monitored in the experiments of Ref. [15] can be derived from the analysis of the Δx profile: the fraction of the translocated chain at which the pore-obstruction event takes place, x^*, the elapsed time at which it occurs, t^*, and the temporal duration of the event, Δt^*. The probability distribution of x^*, t^* and Δt^* are shown in Fig. 5.7a-c.
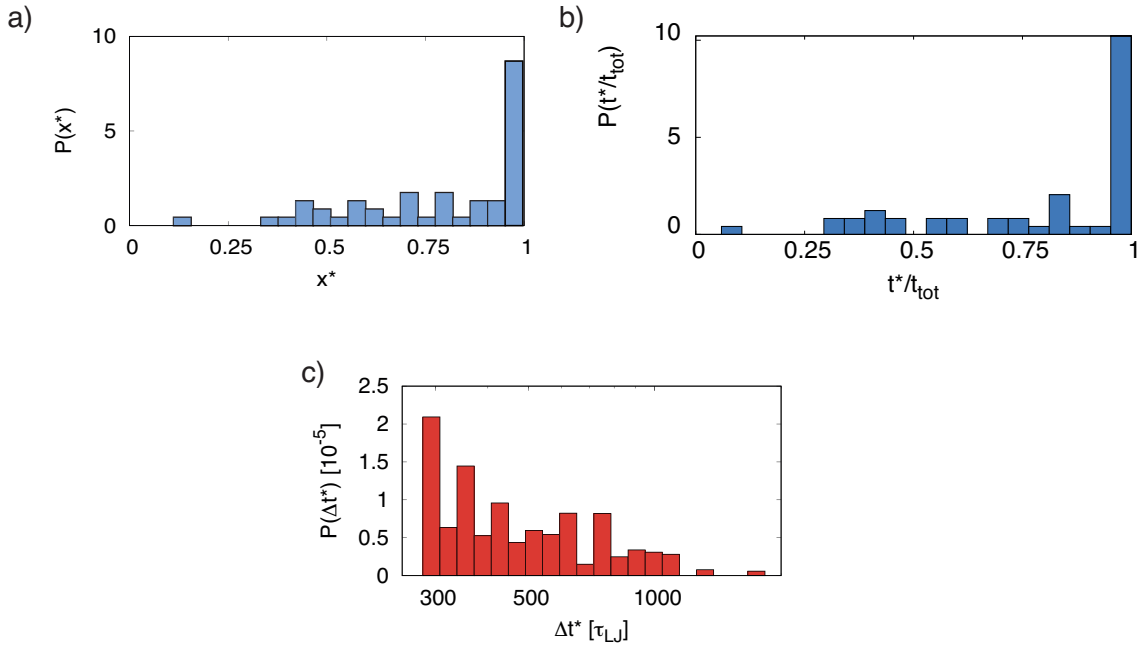
FIGURE 5.7: a) Probability distribution of the translocated chain fraction at the passage (pore obstruction) event, x^*. b) Probability distribution of the normalised time of knot passage, t^*/t_{tot}. t^* is the time of occurrence of the pore obstruction event and t_{tot} is the total translocation time. Given the monotonic relationship between t^* and x^*, this distribution complements the one in a). c) Probability distributions of the event duration of the translocation process, Δt^*.

The key features are two. First, the distribution of x^* (Fig. 5.7a) is skewed towards large values of x^*, and the same holds for the distribution of t^* (Fig. 5.7b), because t^* and x^* are monotonically related. In fact, passage events are virtually absent for $x^* < 0.3$ and the distribution is prominently peaked at $x^* \sim 1$. Secondly, the distribution of the obstruction duration Δt^* (Fig. 5.7c) has an overall decreasing trend, with the shortest obstruction events (which have a minimum duration of 300 τ_{LJ}) being the most probable too. Both these features match the ones reported by Plesa *et al.*, see Fig. 5.1.

This consistency of the experimental and theoretical distributions for x^* and Δt^* is noteworthy given the different contour lengths considered here (10Kbp) and in the experiment (20Kbp or longer). This underscores the robustness of the effects addressed with either of the two approaches. The agreement also gives confidence for using the model to gain insight into aspects that cannot be directly accessed with current experiments. These primarily include various properties of the knotted region, which we discuss in the following.

5.3.3   Knot translocation modes

As we discuss, both the position of the knot along the chain contour and its size affect the Δt^* and x^* distributions in ways that are much richer than previously suspected.

A particularly intriguing relationship is found between the pore obstruction duration, Δt^*, and the size of the knotted region, l_k, when it arrives at the pore. We recall that, as customary, the knotted portion is identified as the shortest portion of the chain that, upon closure, has the same topology of the entire ring. A scatter plot of the two quantities is presented in Fig. 5.8a, where two relevant features are noted.
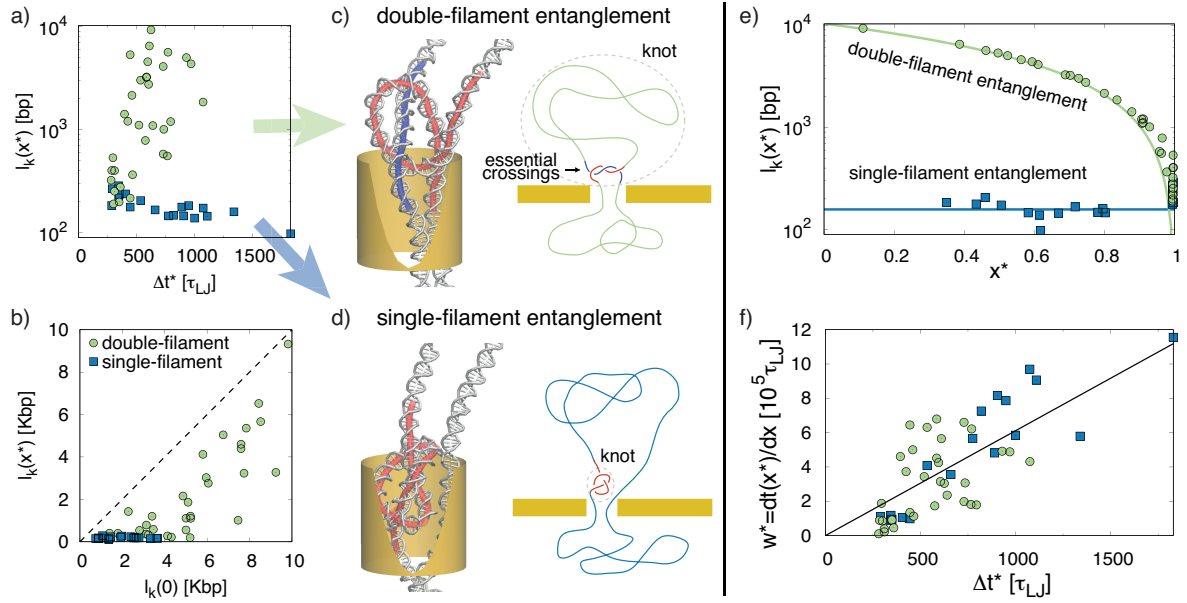
FIGURE 5.8: a) Scatter plot of the knot length at the pore obstruction event, $l_k(x^*)$, versus the duration of the event itself, Δt^*. b) Knot length at the beginning of the translocation process, $l_k(0)$, and at the pore obstruction event, $l_k(x^*)$. Data points are divided in two classes based on whether the knot at the *beginning* of the translocation straddled (green) or not straddled (blue) the site antipodal (on the ring contour) to the translocation initiation site, see main text. For the first group, the pore obstruction event practically involves only the essential crossings of the knotted region, which spans both translocating filaments, see panel c. For the second group, the pore obstruction is caused by a single-filament knot, see panel d. e) Knot length l_k during pore obstruction at x^*. Data points for double-filament knots follow closely the curve $l_k = l_{chain}(1 - x^*)$ (solid green line), while for the other points l_k is about constant and equal to 160bp. f) Scatter plot of Δt^* against the waiting time w^* at the time of passage.

First, the datapoints occupy an L-shaped region. Secondly, for either arm of this region the correlation between Δt^* and l_k is rather weak. Both aspects are not intuitive and, in fact, had not been previously predicted nor envisaged.

The analysis of the trajectories showed that the distinct arms in the diagram of Fig. 5.8a originate from two different modes of knot translocation, as described below.

In the first mode the knot is tight and localised on one of the two translocating filaments, see Fig. 5.8d. This is the most intuitive type of knot passage and, in fact, it was the mode of choice used in Ref. [15] to interpret the experimental data on Δt^* and thus obtain a mapping between pore-obstruction time and knot length. By using a linear mapping, Plesa *et al.* were able to conclude that knots could be rather tight upon translocation, spanning an arclength of tens of nanometers, hence comparable to the DNA persistence length. This result was further put in the context of the elegant theory of metastable knots, which predicts knot localization based on the fact that, in the otherwise broad distribution of knot lengths, the most probable one is about constant - rather than growing - with chain length.

Our results vividly confirm the significant occurrence of tight knots. Indeed, one

observes that the average knot length at the passage event is about 54nm, which is in full accord with the estimate of Plesa *et al.*. This knot length is reached independently of the initial one thanks to tightening of the knot caused by the propagating chain tension, see Fig. 5.8b. One also notes that the l_k versus Δt^* profile in Fig. 5.8a is rather flat for this translocation mode, and hence is different from the linear relationship expected intuitively. An explanation of this effect will be discussed later.

The second, and new mode is associated to the green points in Fig. 5.8. It involves knots that span a significant portion of the ring, consistent with the theoretical results of Ref. [183] on DNA chains of comparable size, which indicated that the most probable knot length is about 2200bp. In fact, these knots experience significantly less tightening during translocation than those discussed above, see Fig. 5.8b. Intriguingly, these knots are large and yet their pore-obstruction times are not at all dissimilar from the tight knot case discussed before.

This conundrum is solved by considering the actual conformation of such rings when the knot is presented at the pore entrance. A typical configuration is shown in Fig. 5.8c. The accompanying sketch clarifies that the knotted portion now spans the entire *cis* part of the ring. This is quantitatively shown in the semi-log plot of Fig. 5.8e where one observes that for this class of knots, the relative chain fraction occupied by the knot is $l_k/l_{chain} \sim (1 - x^*)$.

However, a significant obstruction of the pore occurs only when the region accommodating the essential crossings passes through it. As seen in the figure, this region is typically small, involving 123bp (42nm) on average.

It is therefore this short, essentially-entangled portion of "double-filament" knots, and not their entire contour lengths, l_k, that is captured by Δt^*.

To our knowledge, the possible occurrence of a second mode of translocation, though rather natural *a posteriori*, has not been considered nor foreseen in previous translocation studies, neither for dsDNA rings experiments, nor for simulations of linear, open chains where it can also occur if translocation starts from inside the knot loop region.

Notice that, because the essentially-entangled region is comparable in size to the tight, single-filament knots, the two modes of translocation cannot be distinguished from the sole analysis of Δt^*. This has direct bearings on the interpretation of experimental data. In fact, it poses the necessity to devise suitable means of discriminating or controlling the incidence of the two modes. In this way one could relate more reliably the measured observables to the spontaneous knotting properties of DNA. Our results suggest that this could be achieved, for instance, by suitably choosing the DNA length. The latter, in fact, affects the balance of the two modes, as we discuss later in connection with Fig. 5.9a.

We conclude the analysis of the data in Fig. 5.8a by discussing the second notable feature, namely the lack of a noticeable correlation between the pore obstruction duration, Δt^*, and knot length, l_k. For the second mode of translocation, it is now clear that no obvious relationship between l_k and Δt^* can be expected, because the l_k is not directly informative for the pore obstruction caused by the essentially-entangled region. The case is different, however, for the first mode, i.e. tight single-filament knots, where a proportionality relationship between knot size and passage time appears plausible and was previously surmised [15].

This point is clarified by the plot of Fig. 5.8f, which shows the relationship between Δt^* and w^*, the inverse translocation velocity at the passage event. The two quantities are visibly correlated for both knot translocation modes. Together with the plot Fig. 5.8a the data clarify that of these two properties relatable the passage

time, knot length and knot translocation velocity, the dominant one is the latter. Notice that, because at passage time the contour lengths of single-filament knots and of double-filament essential crossings spans a limited range, from 120bp to 160bp, one has that Δt^* and w^* have an approximate linear proportionality.

This observation might be harnessed to extract further knot-related properties from Δt^*. Because the average translocation velocity depends on the translocated chain fraction, the observed Δt^* vs w^* correlation should effectively subsume a dependence of Δt^* on the knot position along the chain contour, x^*, which could be recovered with sufficient statistics.

5.3.4 Interpretative model

We now consider the origin of the two different translocation modes and of the skewed distribution of the knot position at passage time, x^*.

Both aspects are best illustrated with the following schematic model of the translocation process. In this purposely simplified scheme we assume that the rooting point where the translocation process initiates is equally likely to lie anywhere on the ring contour. We also assume that tension propagates in the same way along the two ring arms, so that they meet at the antipodal midpoint, that is at the point at half ring contour length from the root.

The main discriminator for the two translocation modes is whether the knot is entirely located on only one of the two arms, or whether it straddles the antipodal midpoint and hence spans both arms, see sketches in Fig. 5.9.

In the former case the sliding of the progressively tensioned ring arm causes the knotted region to tighten towards the distal knot end (i.e. the end that is furthest from the rooting point) while it is dragged to the pore. The tightened single-filament knot will then pass through the pore.

Instead, when the knotted region straddles the antipodal midpoint, the knot will be pulled from both sides and will be dragged towards the pore by both tensioned arms. Such double pulling typically causes the essential crossings to become interlocked, trapping the knot in a moderate degree of tightening. The pore obstruction event is then associated to the passage of the essential crossings.

These two cases are directly associated to the two different translocation modes highlighted in Fig. 5.8. So much so that the two sets in the figure were not assigned from an *a posteriori* supervised inspection of the trajectories, but rather *a priori* based on the aforementioned distinction. In fact, the two sets precisely correspond to knots that span a single or both ring arms at the *beginning* of translocation. The neat separation of the two sets of points in Figs. 5.8a,b,e,f supports the viability and usefulness of such discriminatory criterion.

The same criterion can be also used to estimate how the relative incidence of single- versus double-filament entanglement varies with ring length. We considered an ensemble of Monte Carlo-generated rings of length 10, 20 and 50 Kbp, picked a rooting point randomly along their contour and then located the knot on the ring, which we considered open in correspondence of the rooting point. The rings were next assigned to one of the two classes based on whether the knotted region straddled the antipodal midpoint or not. The results, given in Fig. 5.9a, show that the incidence of single-filament entanglement increases steadily with chain length, and goes from 35% for 10Kbp to 70% for 50Kbp. Based on this result, which reflects the interplay of knot and chain lengths [31, 105, 159, 183, 184], we speculate that most of the knot passage events detected in the experiment of Plesa *et al.* pertained to single-stranded entanglements, as implicitly assumed by the authors.
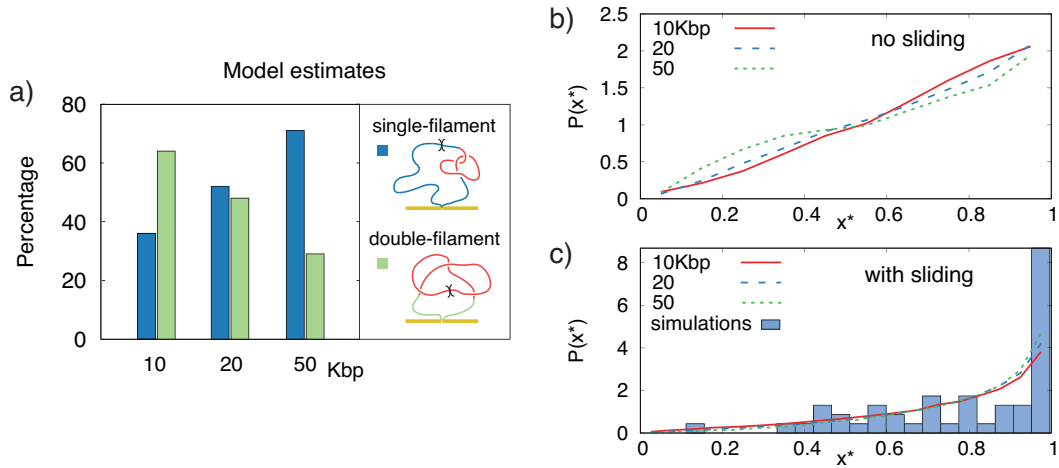
FIGURE 5.9: a) Model estimate of relative percentage of single- and double-filament knots in DNA rings of different length. The estimate considers the length and positioning of the knotted region (highlighted in red in the sketches) with respect to the point (marked with a cross) antipodal to the root. b) The same model is used to predict the translocated chain fraction at the time of the pore obstruction event, x^*, see panel (b). Accounting for the sliding of the knot along the chain brings the model distribution in good quantitative agreement with the actual simulation data, see panel (c).

The same schematic framework can account for the qualitative features of the distribution of x^*, the knot positioning at time of translocation. We considered, again, the Monte Carlo generated ensemble of rings for which we stochastically-picked the rooting points. Next, for double-filament knots (straddling the antipodal midpoint) we picked x^* uniformly between the two knot ends. For the other, single-filament knots, we picked x^* as the distal end of the knot, the one further from the pore. These criteria embody in the simplest possible way the phenomenology described in the previous paragraphs.

The resulting probability distributions for x^*, shown in Fig. 5.9b, are in qualitative agreement with simulation and experimental data. It is seen that, at all ring lengths, the distributions are skewed towards $x^* = 1$. As for the balance of single- and double-filament knots, the skewness too depends on the interplay of knot and ring length. Indeed, the x^* distribution becomes flatter for longer rings, where the knotted region occupies a smaller fraction of the chain contour.

The above modelling scheme neglects the possibility that tight knots may slide on the filament contour. As was clarified in the theoretical study of Ref. [58, 142, 144], such sliding can occur for fully-flexible chains and, in fact, make it possible for individual knotted filaments to fully translocate through very narrow pores, as long as the driving force is not high enough to cause jamming [58, 185, 186]. As a matter of fact, we observed the same knot sliding phenomenology for the present dsDNA system too.

To account for such sliding effect for single-filament knots, we accordingly adjusted the model. Specifically, we assumed that x^* could fall with equal probability between the distal knot end and the antipodal midpoint. The x^* probability distribution predicted by such model is shown in Fig. 5.9c. It presents a noticeably stronger shift forwards x^* values that follows closely the data from the simulated

trajectories. This good level of agreement is somewhat surprising given the simplicity of the model, which does not account for frictional effects related to pore size and force magnitude. Yet, the good accord further corroborates the relevance of sliding effects for dsDNA. We believe, this would be an important avenue to explore further, especially by seeking a quantitative comparison against experimental data. For this, it would probably become essential to take into account the finite resolution of time measurements which could account for the observed effective dependence of the distribution of t^* (related monotonically, but non-linearly, to x^*) on the driving force.

5.4 Conclusions

It is only very recently, that innovative single-molecule techniques have made it possible to detect knots in double-stranded DNA chains driven through nanopores [15]. On the one hand, this gave a striking demonstration of spontaneous knot formation in linear and circularised DNA. On the other hand it also helped unveil a rich and complex phenomenology that, though expectedly relevant for the *in vivo* processing of DNA filaments, is still largely unexplored.

Here, to advance the understanding and characterization of such phenomenology, we studied theoretically the pore translocation of knotted DNA rings using an accurate coarse-grained model for DNA and stochastic molecular dynamics simulations.

We find good agreement with the experimental data, particularly regarding the remarkably brief duration of pore-obstruction events associated to the passage of the knot. By profiling the dynamical evolution of the knotted DNA rings we expose unexpectedly rich properties of the process that cannot be directly accessed in current experiments.

First, we found that translocation of the knotted region can occur in two qualitatively different modes depending on whether the knot is dragged to the pore by only one of the ring arms, or both. In the latter case, knots are typically not tight, and yet we find that the pore obstruction time can be small (as in experiments) because the essential crossings of the knot coalesce in a short region.

Secondly, we found that the sliding and tensioning of the translocating knot causes the same bias towards late knot passage events found in experiments, which had remained unexplained. We finally showed that one of the key determinants of the pore obstruction duration is the initial positioning of the knot along the chain, and suggest how this effect might be deconvolved in experimental measurements for a more precise determination of the length of the region accommodating a knot or its essential crossings. In particular, the occurring phenomenon of knot sliding might give an important contribution. This might be exploitable in future experiments, along with chain length and pore size variations, to discriminate the two modes. Further relevant avenues include the impact on pore translocation of complex topologies such as composite knots, which have so far been characterised for flexible chains only [58], as well as the geometry-topology interplay in DNA rings that cannot relax supercoiling and torsional stress [187].

This first theoretical account, provides a detailed and physically-appealing insight into phenomenology of knotted dsDNA pore translocation. It provides a valuable and transparent interpretative framework for available experimental data while pointing out specific directions for new experiments as well as theoretical ones aimed

at better understanding the implication of intra-chain entanglement for the *in vivo* processing of DNA, and possibly other biopolymers too.

Bibliography

[1]   B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, New York, 2002.

[2]   H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, J. Darnell, et al. *Molecular cell biology*. Scientific American Books New York, 1995.

[3]   H. L. Frisch and E. Wasserman. "Chemical topology". In: *J. Am. Chem. Soc.* 83.18 (1961), pp. 3789–3795.

[4]   M. Delbruck. *Knotting problems in biology, in: Mathematical problems in the biological sciences*. Vol. 14. American Mathematical Soc., 1962, pp. 55–63.

[5]   D. W. Sumners and S. G. Whittington. "Knots in self-avoiding walks". In: *J. Phys. A* 21 (1988), pp. 1689–1694.

[6]   E. J. Van Rensburg, D. Sumners, E Wasserman, and S. Whittington. "Entanglement complexity of self-avoiding walks". In: *Journal of Physics A: Math. Gen.* 25.24 (1992), p. 6557.

[7]   H. Wang, R. J. Di Gate, and N. C. Seeman. "An RNA topoisomerase". In: *Proc. Natl. Acad. Sci. U.S.A.* 93.18 (1996), pp. 9477–9482.

[8]   J. S. Richardson. "β-Sheet topology and the relatedness of proteins". In: *Nature* 268 (1977), pp. 495–500.

[9]   M. L. Mansfield. "Are there knots in proteins?" In: *Nat. Struct. Mol. Biol.* 1.4 (1994), pp. 213–214.

[10] W. R. Taylor. "A deeply knotted protein structure and how it might fold". In: *Nature* 406.6798 (2000), pp. 916–919.

[11] L. F. Liu, R. E. Depew, and J. C. Wang. "Knotted single-stranded DNA rings: A novel topological isomer of circular single-stranded DNA formed by treatment with Escherichia coli ω protein". In: *J. Mol. Biol.* 106.2 (1976), pp. 439–452.

[12] V. V. Rybenkov, N. R. Cozzarelli, and A. V. Vologodskii. "Probability of DNA knotting and the effective diameter of the DNA double helix". In: *Proc. Natl. Acad. Sci. U.S.A.* 90.11 (1993), pp. 5307–5311.

[13] S. Wasserman and N. R. Cozzarelli. "Biochemical topology: applications to DNA recombination and replication". In: *Science* 232 (1986), pp. 951–961.

[14] J. Arsuaga, M. Vazquez, P. McGuirk, S. Trigueros, J. Roca, et al. "DNA knots reveal a chiral organization of DNA in phage capsids". In: *Proc. Natl. Acad. Sci. USA* 102.26 (2005), pp. 9165–9169.

[15] C. Plesa, D. Verschueren, S. Pud, J. van der Torre, J. W. Ruitenberg, M. J. Witteveen, M. P. Jonsson, A. Y. Grosberg, Y. Rabin, and C. Dekker. "Direct observation of DNA knots using a solid-state nanopore". In: *Nat. Nanotechnol.* 11 (2016), pp. 1093–1097.

[16] C. Micheletti, M. Di Stefano, and H. Orland. "Absence of knots in known RNA structures". In: *Proc. Natl. Acad. Sci. U.S.A.* 112.7 (2015), pp. 2052–2057.

[17]   P. Virnau, L. A. Mirny, and M. Kardar. "Intricate knots in proteins: Function and evolution". In: *PLOS Comput. Biol.* 2.9 (2006), e122.

[18]   R. W. Deibler, J. K. Mann, L. S. De Witt, and L. Zechiedrich. "Hin-mediated DNA knotting and recombining promote replicon dysfunction and mutation". In: *BMC molecular biology* 8.1 (2007), p. 44.

[19]   K Reidemeister. *Knotentheorie, Ergebnisse der Matematik und ihrer Grenzgebiete.* Vol. Band 1, Heft 1. Springer Berlin, 1932.

[20]   D. Rolfsen. *Knots and links.* Vol. 346. American Mathematical Soc., 1976.

[21]   W. R. Lickorish. *An introduction to knot theory.* Vol. 175 of Graduate Texts in Mathematics. Springer-Verlag, New York, 1997.

[22]   C. Ernst and D. Sumners. "The growth of the number of prime knots". In: *Mathematical Proceedings of the Cambridge Philosophical Society.* Vol. 102. 02. Cambridge Univ. Press. 1987, pp. 303–315.

[23]   D. J. Welsh. *On the number of knots and links.* Rhenische Friedrich-Wilhelms-Universität Bonn, 1991.

[24]   J. Hoste, M. Thistlethwaite, and J. Weeks. "The first 1,701,936 knots". In: *The Mathematical Intelligencer* 20.4 (1998), pp. 33–48.

[25]   J. W. Alexander. "Topological invariants of knots and links". In: *Trans. Amer. Math. Soc.* 30.2 (1928), pp. 275–306.

[26]   E. Orlandini and S. G. Whittington. "Statistical topology of closed curves: some applications in polymer physics". In: *Rev. Mod. Phys.* 79 (2007), p. 611.

[27]   C. Micheletti, D. Marenduzzo, and E. Orlandini. "Polymers with spatial or topological constraints: Theoretical and computational results". In: *Phys. Rep.* 504 (2011), p. 1.

[28]   A. Stasiak, A. Dobay, J. Dubochet, and G. Dietler. "Knotted fishing line, covalent bonds, and breaking points". In: *Science* 286.5437 (1999), pp. 11–11.

[29]   L. Tubiana, E. Orlandini, and C. Micheletti. "Probing the entanglement and locating knots in ring polymers: a comparative study of different arc closure schemes". In: *Prog. Theor. Phys.* 191 (2011), pp. 192–204.

[30]   V. Katritch, W. K. Olson, A. Vologodskii, J. Dubochet, and A. Stasiak. "Tightness of random knotting". In: *Phys. Rev. E* 61 (2000), p. 5545.

[31]   B Marcone, E Orlandini, A. Stella, and F Zonta. "What is the length of a knot in a polymer?" In: *Journal of Physics A: Math. Gen.* 38.1 (2004), p. L15.

[32]   M. L. Mansfield and J. F. Douglas. "Properties of knotted ring polymers. I. Equilibrium dimensions". In: *J. Chem. Phys.* 133 (2010), p. 044903.

[33]   P. Virnau, A. Mallam, and S. Jackson. "Structures and folding pathways of topologically knotted proteins." In: *J. Phys. Condens. matter* 23.3 (2011), p. 033101.

[34]   A. L. Mallam. "How does a knotted protein fold?" In: *Febs Journal* 276.2 (2009), pp. 365–375.

[35]   T. O. Yeates, T. S. Norcross, and N. P. King. "Knotted and topologically complex proteins as models for studying folding and stability". In: *Curr. Opin. Chem. Biol.* 11.6 (2007), pp. 595–603.

[36]   P. F. Faísca. "Knotted proteins: A tangled tale of Structural Biology". In: *Comput. Struct. Biotechnol. J.* 13 (2015), pp. 459–468.

[37]   R. Potestio, C. Micheletti, and H. Orland. "Knotted vs. unknotted proteins: evidence of knot-promoting loops". In: *PLOS Comput. Biol.* 6.7 (2010), e1000864.

[38]   S. E. Jackson, A. Suma, and C. Micheletti. "How to fold intricately: using theory and experiments to unravel the properties of knotted proteins". In: *Curr. Opin. Struc. Biol.* 42 (2017), pp. 6–14.

[39]   S. Chakraborty, R. Venkatramani, B. J. Rao, B. Asgeirsson, and A. M. Dandekar. "Protein structure quality assessment based on the distance profiles of consecutive backbone C_α atoms". In: *F1000Res.* 2 (2013).

[40]   S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. "Basic local alignment search tool". In: *J. Mol. Biol.* 215.3 (1990), pp. 403–410.

[41]   Y. Huang, B. Niu, Y. Gao, L. Fu, and W. Li. "CD-HIT Suite: a web server for clustering and comparing biological sequences". In: *Bioinformatics* 26.5 (2010), pp. 680–682.

[42]   W. R. Taylor. "Protein knots and fold complexity: some new twists". In: *Comput. Biol. Chem.* 31.3 (2007), pp. 151–162.

[43]   N. C. Lim and S. E. Jackson. "Molecular knots in biology and chemistry". In: *J. Phys. Condens. Matter* 27.35 (2015), p. 354101.

[44]   R. C. Lua and A. Y. Grosberg. "Statistics of knots, geometry of conformations, and evolution of proteins". In: *PLOS Comput. Biol.* 2.5 (2006), e45.

[45]   P. Virnau, Y. Kantor, and M. Kardar. "Knots in globule and coil phases of a model polyethylene". In: *J. Am. Chem. Soc.* 127 (2005), p. 15102.

[46]   L. Tubiana, A. Rosa, F. Fragiacomo, and C. Micheletti. "Spontaneous knotting and unknotting of flexible linear polymers: Equilibrium and kinetic aspects". In: *Macromolecules* 46.9 (2013), pp. 3669–3678.

[47]   A. L. Mallam and S. E. Jackson. "Folding studies on a knotted protein". In: *J. Mol. Biol.* 346.5 (2005), pp. 1409–1421.

[48]   A. L. Mallam and S. E. Jackson. "Probing nature's knots: the folding pathway of a knotted homodimeric protein". In: *J. Mol. Biol.* 359.5 (2006), pp. 1420–1436.

[49]   A. L. Mallam, E. R. Morris, and S. E. Jackson. "Exploring knotting mechanisms in protein folding". In: *Proc. Natl. Acad. Sci. U.S.A.* 105.48 (2008), pp. 18740–18745.

[50]   A. L. Mallam, J. M. Rogers, and S. E. Jackson. "Experimental detection of knotted conformations in denatured proteins". In: *Proc. Natl. Acad. Sci. U.S.A.* 107.18 (2010), pp. 8189–8194.

[51]   N. C. Lim and S. E. Jackson. "Mechanistic insights into the folding of knotted proteins in vitro and in vivo". In: *J. Mol. Biol.* 427.2 (2015), pp. 248–258.

[52]   E. Leroy, R. Boyer, G. Auburger, B. Leube, G. Ulm, E. Mezey, G. Harta, M. J. Brownstein, S. Jonnalagada, T. Chernova, et al. "The ubiquitin pathway in Parkinson's disease". In: *Nature* 395.6701 (1998), p. 451.

[53]   Y. Liu, L. Fallon, H. A. Lashuel, Z. Liu, and P. T. Lansbury. "The UCH-L1 gene encodes two opposing enzymatic activities that affect α-synuclein degradation and Parkinson's disease susceptibility". In: *Cell* 111.2 (2002), pp. 209–218.

[54] F. I. Andersson, E. F. Werrell, L. McMorran, W. J. Crone, C. Das, S.-T. D. Hsu, and S. E. Jackson. "The effect of Parkinson's-disease-associated mutations on the deubiquitinating enzyme UCH-L1". In: *J. Mol. Biol.* 407.2 (2011), pp. 261–272.

[55] S.-C. Lou, S. Wetzel, H. Zhang, E. W. Crone, Y.-T. Lee, S. E. Jackson, and S.-T. D. Hsu. "The knotted protein UCH-L1 exhibits partially unfolded forms under native conditions that share common structural features with its kinetic folding intermediates". In: *J. Mol. Biol.* 428.11 (2016), pp. 2507–2520.

[56] F. I. Andersson, D. G. Pina, A. L. Mallam, G. Blaser, and S. E. Jackson. "Untangling the folding mechanism of the 52-knotted protein UCH-L3". In: *Febs Journal* 276.9 (2009), pp. 2625–2635.

[57] M. Wojciechowski, À. Gómez-Sicilia, M. Carrión-Vázquez, and M. Cieplak. "Unfolding knots by proteasome-like systems: simulations of the behaviour of folded and neurotoxic proteins". In: *Mol. Biosyst.* 12.9 (2016), pp. 2700–2712.

[58] A. Suma, A. Rosa, and C. Micheletti. "Pore Translocation of Knotted Polymer Chains: How Friction Depends on Knot Complexity". In: *ACS Macro Lett.* 4.12 (2015), pp. 1420–1424.

[59] A. Suma, E. Orlandini, and C. Micheletti. "Knotting dynamics of DNA chains of different length confined in nanochannels". In: *J. Phys. Condens. Matter* 27.35 (2015), p. 354102.

[60] S. Y. Shaw and J. C. Wang. "Knotting of a DNA chain during ring closure". In: *Science* 260.5107 (1993), pp. 533–537.

[61] L. F. Liu, C.-C. Liu, and B. M. Alberts. "Type II DNA topoisomerases: enzymes that can unknot a topologically knotted DNA molecule via a reversible double-strand break". In: *Cell* 19.3 (1980), pp. 697–707.

[62] V. V. Rybenkov, N. R. Cozzarelli, and A. V. Vologodskii. "Probability of DNA knotting and the effective diameter of the DNA double helix". In: *Proc. Natl. Acad. Sci. USA* 90 (1993), p. 5307.

[63] J. Arsuaga, M. Vázquez, S. Trigueros, J. Roca, et al. "Knotting probability of DNA molecules confined in restricted volumes: DNA knotting in phage capsids". In: *Proc. Natl. Acad. Sci. USA* 99.8 (2002), pp. 5373–5377.

[64] E. Ercolini, F. Valle, J. Adamcik, G. Witz, R. Metzler, P. De Los Rios, J. Roca, and G. Dietler. "Fractal Dimension and Localization of DNA Knots". In: *Phys. Rev. Lett.* 98 (2007), p. 058102.

[65] K. Shishido, N. Komiyama, and S. Ikawa. "Increased production of a knotted form of plasmid pBR322 DNA in Escherichia coli DNA topoisomerase mutants". In: *J. Mol. Biol.* 195.1 (1987), pp. 215–218.

[66] T. Goto and J. C. Wang. "Yeast DNA topoisomerase II. An ATP-dependent type II topoisomerase that catalyzes the catenation, decatenation, unknotting, and relaxation of double-stranded DNA rings." In: *Journal of Biological Chemistry* 257.10 (1982), pp. 5866–5872.

[67] J. Arsuaga, M. Vázquez, S. Trigueros, J. Roca, et al. "Knotting probability of DNA molecules confined in restricted volumes: DNA knotting in phage capsids". In: *Proc. Natl. Acad. Sci. USA* 99 (2002), p. 5373.

[68] D. Marenduzzo, C. Micheletti, E. Orlandini, and D. W. Sumners. "Topological friction strongly affects viral DNA ejection." In: *Proc. Natl. Acad. Sci. USA* 110.50 (2013), pp. 20081–20086.

[69] D. Marenduzzo, E. Orlandini, A. Stasiak, L. Tubiana, C. Micheletti, et al. "DNA–DNA interactions in bacteriophage capsids are responsible for the observed DNA knotting". In: *Proc. Natl. Acad. Sci. USA* 106.52 (2009), pp. 22269–22274.

[70] J. O. Tegenfeldt, C. Prinz, H. Cao, S. Chou, W. W. Reisner, R. Riehn, Y. M. Wang, E. C. Cox, J. C. Sturm, P. Silberzan, and R. H. Austin. "From the Cover: The dynamics of genomic-length DNA molecules in 100-nm channels." In: *Proc. Natl. Acad. Sci. U.S.A.* 101.30 (2004), pp. 10979–10983.

[71] D. Stein, F. H. J. van der Heyden, W. J. A. Koopmans, and C. Dekker. "Pressure-driven transport of confined DNA polymers in fluidic channels". In: *Proc. Natl. Acad. Sci. USA* 103 (2006), pp. 15853–15858.

[72] M. Muthukumar. "Mechanism of DNA transport through pores." In: *Annu. Rev. Biophys. Biomol. Struct.* 36 (2007), pp. 435–450.

[73] F. Persson, P. Utko, W. Reisner, N. B. Larsen, and A. Kristensen. "Confinement spectroscopy: probing single DNA molecules with tapered nanochannels". In: *Nano Lett.* 9.4 (2009), pp. 1382–1385.

[74] P Cifra. "Channel confinement of flexible and semiflexible macromolecules". In: *J. Chem. Phys.* 131.22 (2009), pp. 224903–224903.

[75] D. R. Tree, Y. Wang, and K. D. Dorfman. "Modeling the relaxation time of DNA confined in a nanochannel." In: *Biomicrofluidics* 7.5 (2013), p. 54118.

[76] W. Reisner, K. J. Morton, R. Riehn, Y. M. Wang, Z. Yu, M. Rosen, J. C. Sturm, S. Y. Chou, E. Frey, and R. H. Austin. "Statics and dynamics of single DNA molecules confined in nanochannels". In: *Phys. Rev. Lett.* 94 (2005), p. 196101.

[77] Y.-L. Chen, M. Graham, J. de Pablo, G. Randall, M Gupta, and P. Doyle. "Conformation and dynamics of single DNA molecules in parallel-plate slit microchannels". In: *Phys. Rev. E* 70.6 (2004), p. 060901.

[78] D. J. Bonthuis, C. Meyer, D. Stein, and C. Dekker. "Conformation and Dynamics of DNA Confined in Slitlike Nanofluidic Channels". In: *Phys. Rev. Lett.* 101 (2008), p. 108303.

[79] P.-G. de Gennes. *Scaling concepts in Polymer Physics*. Cornell University Press, Ithaca, New York, 1979.

[80] T. Odijk. "The statistics and dynamics of confined or entangled stiff polymers". In: *Macromolecules* 16 (1983), pp. 1340–1344.

[81] L. Dai, S. Y. Ng, P. S. Doyle, and J. R. C. van der Maarel. "Conformation Model of Back-Folding and Looping of a Single DNA Molecule Confined Inside a Nanochannel". In: *ACS Macro Lett.* 1 (2012), pp. 1046–1050.

[82] A. Muralidhar, D. Tree, Y. Wang, and K. Dorfman. "Interplay between chain stiffness and excluded volume of semiflexible polymers confined in nanochannels". In: *J. Chem. Phys.* 140 (2014), p. 084905.

[83] Y. Wang, D. Tree, and K. Dorfman. "Simulation of DNA Extension in Nanochannels." In: *Macromolecules* 44.16 (2011), pp. 6594–6604.

[84] T. Odijk. "Scaling theory of DNA confined in nanochannels and nanoslits". In: *Phys. Rev. E* 77.6 (2008), p. 060901.

[85]   Y. L. Chen. "Electro-entropic excluded volume effects on DNA looping and relaxation in nanochannels". In: *Biomicrofluidics* 7 (2013), p. 054119.

[86]   A. Muralidhar, D. R. Tree, and K. D. Dorfman. "Backfolding of Wormlike Chains Confined in Nanochannels". In: *Macromolecules* (2014).

[87]   C. Micheletti and E. Orlandini. "Knotting and metric scaling properties of DNA confined in nano-channels: a Monte Carlo study". In: *Soft Matter* 8 (2012), pp. 10959–10968.

[88]   L. Dai, C. B. Renner, and P. S. Doyle. "Metastable knots in confined semiflexible chains". In: *Macromolecules* 48.8 (2015), pp. 2812–2818.

[89]   W. Mobius, E. Frey, and U. Gerland. "Spontaneous unknotting of a polymer confined in a nanochannel." In: *Nano Lett.* (2008), pp. 4518–4522.

[90]   A. Jain and K. D. Dorfman. "Simulations of knotting of DNA during genome mapping". In: *Biomicrofluidics* 11.2 (2017), p. 024117.

[91]   E. T. Lam, A. Hastie, C. Lin, D. Ehrlich, S. K. Das, M. D. Austin, P. Deshpande, H. Cao, N. Nagarajan, M. Xiao, and P.-Y. Kwok. "Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly." In: *Nat. Biotechnol.* 30.8 (2012), pp. 771–776.

[92]   M. Levy-Sakin and Y. Ebenstein. "Beyond sequencing: optical mapping of DNA in the age of nanotechnology and nanoscopy". In: *Curr. Opin. Biotechnol.* 24.4 (2013), pp. 690–698.

[93]   W. Reisner, J. N. Pedersen, and R. H. Austin. "DNA confinement in nanochannels: physics and biological applications". In: *Rep. Prog. Phys.* 75 (2012), p. 106601.

[94]   K. Dorfman, S. B. W. King, D. W. Olson, and D. Thomas J. D. R.and Tree. "Beyond Gel Electrophoresis: Microfluidic Separations, Fluorescence Burst Analysis, and DNA Stretching". In: *Chem. Rev.* 113.4 (2013), 2584¿2667.

[95]   C. Micheletti and E. Orlandini. "Knotting and Unknotting Dynamics of DNA Strands in Nanochannels". In: *ACS Macro Lett.* 3.9 (2014), pp. 876–880.

[96]   K. Kremer and G. S. Grest. "Dynamics of entangled linear polymer melts: A molecular-dynamics simulation". In: *J. Chem. Phys.* 92.8 (1990), pp. 5057–5086.

[97]   S. Plimpton. "Fast parallel algorithms for short-range molecular dynamics". In: *J. Comput. Phys.* 117.1 (1995), pp. 1–19.

[98]   B Marcone, E Orlandini, A. Stella, and F Zonta. "Size of knots in ring polymers". In: *Phys. Rev. E* 75.4 (2007), p. 041105.

[99]   L. Tubiana, E. Orlandini, and C. Micheletti. "Multiscale Entanglement in Ring Polymers under Spherical Confinement". In: *Phys. Rev. Lett.* 107 (2011), p. 188302.

[100] X. R. Bao, H. J. Lee, and S. R. Quake. "Behavior of Complex Knots in Single DNA Molecules". In: *Phys. Rev. Lett.* 91 (2003), p. 265506.

[101] A. Vologodskii. "Brownian Dynamics Simulation of Knot Diffusion along a Stretched {DNA} Molecule". In: *Biophys. J.* 90.5 (2006), pp. 1594 –1597.

[102] L. Huang and D. E. Makarov. "Langevin Dynamics Simulations of the Diffusion of Molecular Knots in Tensioned Polymer Chains". In: *J. Phys. Chem. A* 111.41 (2007), pp. 10338–10344.

[103] M. Di Stefano, L. Tubiana, M. Di Ventra, and C. Micheletti. "Driving knots on DNA with AC/DC electric fields: topological friction and memory effects." In: *Soft Matter* 10.34 (2014), pp. 6491–6498.

[104]   R. Metzler, W. Reisner, R. Riehn, R. Austin, J. O. Tegenfeldt, and I. M. Sokolov. "Diffusion mechanisms of localised knots along a polymer". In: *EPL* 76.4 (2006), p. 696.

[105]   P. G. Dommersnes, Y. Kantor, and M. Kardar. "Knots in charged polymers". In: *Phys. Rev. E* 66.3 (2002), p. 031802.

[106]   S. Redner. *A guide to first-passage processes*. Cambridge University Press, 2001.

[107]   M. Zwolak and M. D. Ventra. "*Colloquium*: Physical approaches to DNA sequencing and detection". In: *Rev. Mod. Phys.* 80 (2008), pp. 141–165.

[108]   J. Chuang, Y. Kantor, and M. Kardar. "Anomalous dynamics of translocation". In: *Phys. Rev. E* 65.1 (2001), p. 011802.

[109]   Y. Kantor and M. Kardar. "Anomalous dynamics of forced translocation". In: *Phys. Rev. E* 69 (2004), p. 021806.

[110]   A. Y. Grosberg, S Nechaev, M Tamm, and O Vasilyev. "How long does it take to pull an ideal polymer into a small hole?" In: *Phys. Rev. Lett.* 96.22 (2006), p. 228105.

[111]   C. Chatelain, Y. Kantor, and M. Kardar. "Probability distributions for polymer translocation". In: *Phys. Rev. E* 78 (2008), p. 021129.

[112]   V. V. Palyulin, T. Ala-Nissila, and R. Metzler. "Polymer translocation: the first two decades and the recent diversification". In: *Soft Matter* 10 (45 2014), pp. 9016–9037.

[113]   W. F. Reinhart, J. G. Reifenberger, D. Gupta, A. Muralidhar, J. Sheats, H. Cao, and K. D. Dorfman. "Distribution of distances between DNA barcode labels in nanochannels close to the persistence length." In: *J. Chem. Phys.* 142.6 (2015), p. 064902.

[114]   L. Huang and D. E. Makarov. "Translocation of a knotted polypeptide through a pore". In: *J. Chem. Phys.* 129.12 (2008), pp. 121107–121107.

[115]   B. Luan and A. Aksimentiev. "Electro-osmotic screening of the DNA charge in a nanopore". In: *Phys. Rev. E* 78 (2 2008), p. 021912.

[116]   C. Plesa, S. W. Kowalczyk, R. Zinsmeester, A. Y. Grosberg, Y. Rabin, and C. Dekker. "Fast translocation of proteins through solid state nanopores". In: *Nano Lett.* 13.2 (2013), pp. 658–663.

[117]   G. F. Schneider and C. Dekker. "DNA sequencing with nanopores". In: *Nat. Nanotechnol.* 30.4 (2012), pp. 326–328.

[118]   K Luo, R Metzler, T Ala-Nissila, and S. C. Ying. "Polymer translocation out of confined environments". In: *Phys. Rev. E* 80.2 (2009), pp. 021907–021907.

[119]   K Luo and R Metzler. "Polymer translocation into a fluidic channel through a nanopore". In: *Phys. Rev. E* 82.2 (2010), pp. 021922–021922.

[120]   K Luo and R Metzler. "The chain sucker: translocation dynamics of a polymer chain into a long narrow channel driven by longitudinal flow". In: *J. Chem. Phys.* 134.13 (2011), pp. 135102–135102.

[121]   D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. Huang, et al. "The potential and challenges of nanopore sequencing". In: *Nat. Biotechnol.* 26 (2008), p. 1146.

[122]   S. Van Dorp, U. F. Keyser, N. H. Dekker, C. Dekker, and S. G. Lemay. "Origin of the electrophoretic force on DNA in solid-state nanopores". In: *Nat. Phys.* 5 (2009), p. 347.

[123] A. Kravats, M. Jayasinghe, and G. Stan. "Unfolding and translocation pathway of substrate protein controlled by structure in repetitive allosteric cycles of the ClpY ATPase." In: *Proc. Natl. Acad. Sci. USA* 108.6 (2011), pp. 2234–2239.

[124] J. Larkin, R. Y. Henley, M. Muthukumar, J. K. Rosenstein, and M. Wanunu. "High-bandwidth protein analysis using solid-state nanopores." In: *Biophys. J.* 106.3 (2014), pp. 696–704.

[125] F Traversi, C Raillon, S. Benameur, K Liu, S Khlybov, M Tosun, D Krasnozhon, A Kis, and A Radenovic. "Detecting the translocation of DNA through a nanopore using graphene nanoribbons". In: *Nat. Nanotechnol.* 8.12 (2013), pp. 939–945.

[126] E. H. Trepagnier, A. Radenovic, D. Sivak, P. Geissler, and J. Liphardt. "Controlling DNA capture and propagation through artificial nanopores". In: *Nano Lett.* 7.9 (2007), pp. 2824–2830.

[127] P. Jing, F. Haque, D. Shu, C. Montemagno, and P. Guo. "One-way traffic of a viral motor channel for double-stranded DNA translocation". In: *Nano Lett.* 10.9 (2010), pp. 3620–3627.

[128] J. Chen, A. Tsai, S. E. O'Leary, A. Petrov, and J. D. Puglisi. "Unraveling the dynamics of ribosome translocation". In: *Curr. Opin. Struc. Biol.* 22.6 (2012), pp. 804–814.

[129] E. Schleiff and T. Becker. "Common ground for protein translocation: access control for mitochondria and chloroplasts". In: *Nat. Rev. Mol. Cell Biol.* 12.1 (2011), pp. 48–59.

[130] D. Deamer, M. Akeson, and D. Branton. "Three decades of nanopore sequencing". In: *Nat. Biotechnol.* 34.5 (2016), pp. 518–524.

[131] S. J. Heerema and C. Dekker. "Graphene nanodevices for DNA sequencing". In: *Nat. Nanotechnol.* 11.2 (2016), pp. 127–136.

[132] S. Agah, M. Zheng, M. Pasquali, and A. B. Kolomeisky. "DNA sequencing by nanopores: advances and challenges". In: *J. Phys. D* 49.41 (2016), p. 413001.

[133] Y. Feng, Y. Zhang, C. Ying, D. Wang, and C. Du. "Nanopore-based fourth-generation DNA sequencing technology". In: *Genomics, proteomics & bioinformatics* 13.1 (2015), pp. 4–16.

[134] L. Steinbock and A Radenovic. "The emergence of nanopores in next-generation sequencing". In: *Nanotechnology* 26.7 (2015), p. 074003.

[135] M. Tsutsui, M. Taniguchi, K. Yokota, and T. Kawai. "Identifying single nucleotides by tunnelling current". In: *Nat. Nanotechnol.* 5.4 (2010), pp. 286–290.

[136] M. Di Ventra and M. Taniguchi. "Decoding DNA, RNA and peptides with quantum tunnelling". In: *Nat. Nanotechnol.* 11.2 (2016), pp. 117–126.

[137] J. Feng, K Liu, R. Bulushev, S Khlybov, D Dumcenco, A Kis, A Radenovic, et al. "Identification of single nucleotides in MoS2 nanopores". In: *Nat. Nanotechnol.* 10.12 (2015), pp. 1070–1070.

[138] J. J. Kasianowicz, E. Brandin, D. Branton, and D. W. Deamer. "Characterization of individual polynucleotide molecules using a membrane channel". In: *Proc. Natl. Acad. Sci. U.S.A.* 93.24 (1996), pp. 13770–13773.

[139] J. Michels and F. Wiegel. "The distribution of the Alexander polynomials of knots confined to a thin layer". In: *J. Phys. A* 22 (1989), pp. 2393–2398.

[140] M. K. Shimamura and T. Deguchi. "Knot complexity and the probability of random knotting." In: *Phys. Rev. E* 66.4 (2002), p. 040801.

[141] R. Matthews, A. A. Louis, and J. M. Yeomans. "Knot-Controlled Ejection of a Polymer from a Virus Capsid". In: *Phys. Rev. Lett.* 102 (2009), p. 088101.

[142] A. Rosa, M. Di Ventra, and C. Micheletti. "Topological jamming of spontaneously knotted polyelectrolyte chains driven through a nanopore." In: *Phys. Rev. Lett.* 109.11 (2012), p. 118301.

[143] P Szymczak. "Translocation of knotted proteins through a pore". In: *Eur. Phys. J. Special Topics* 223.9 (2014), pp. 1805–1812.

[144] V. Narsimhan, C. B. Renner, and P. S. Doyle. "Translocation dynamics of knotted polymers under a constant or periodic external field". In: *Soft matter* 12.22 (2016), pp. 5041–5049.

[145] F. C. Rieger and P. Virnau. "A Monte Carlo study of knots in long double-stranded DNA chains". In: *PLOS Comput. Biol.* 12.9 (2016), e1005029.

[146] M. Doi and S. F. Edwards. *The Theory of Polymer Dynamics*. New York: Oxford University Press, 1986.

[147] J. Lagerqvist, M. Zwolak, and M. Di Ventra. "Fast DNA Sequencing via Transverse Electronic Transport". In: *Nano Lett.* 6.4 (2006), pp. 779–782.

[148] T. Ohshiro, K. Matsubara, M. Tsutsui, M. Furuhashi, M. Taniguchi, and T. Kawai. "Single-molecule electrical random resequencing of DNA and RNA." In: *Sci. Rep.* 2 (2012), p. 501.

[149] Y. He, M. Tsutsui, R. H. Scheicher, F. Bai, M. Taniguchi, and T. Kawai. "Thermophoretic manipulation of DNA translocation through nanopores." In: *ACS nano* 7.1 (2013), pp. 538–546.

[150] S. Rahong, T. Yasui, T. Yanagida, K. Nagashima, M. Kanai, G. Meng, Y. He, F. Zhuge, N. Kaji, T. Kawai, and Y. Baba. "Self-assembled nanowire arrays as three-dimensional nanopores for filtration of DNA molecules." In: *An. Sci.* 31.3 (2015), pp. 153–157.

[151] C. Maffeo, R. Schöpflin, H. Brutzer, R. Stehr, A. Aksimentiev, G. Wedemann, and R. Seidel. "DNA-DNA interactions in tight supercoils are described by a small effective charge density." In: *Phys. Rev. Lett.* 105.15 (2010), p. 158101.

[152] I. M. Derrington, T. Z. Butler, M. D. Collins, E. Manrao, M. Pavlenok, M. Niederweis, and J. H. Gundlach. "Nanopore DNA sequencing with MspA". In: *Proc. Natl. Acad. Sci. USA* 107 (2010), p. 16060.

[153] R. Matthews, A. A. Louis, and J. M. Yeomans. "Effect of topology on dynamics of knots in polymers under tension". In: *EPL* 89.2 (2010), p. 20001.

[154] C. B. Renner and P. S. Doyle. "Untying Knotted DNA with Elongational Flows". In: *ACS Macro Lett.* 3.10 (2014), pp. 963–967.

[155] B. Trefz, J. Siebert, and P. Virnau. "How molecular knots can pass through each other". In: *Proc. Natl. Acad. Sci. U.S.A.* 111.22 (2014), pp. 7948–7951.

[156] A. Konyukhov and K. Schweizerhof. *Computational Contact Mechanics*. Berlin: Springer, 2013.

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

[158] A. Suma and C. Micheletti. "Pore translocation of knotted DNA rings". In: *Proc. Natl. Acad. Sci. U.S.A.* 114.15 (2017), E2991–E2997.

[159] A. Y. Grosberg and Y. Rabin. "Metastable Tight Knots in a Wormlike Polymer". In: *Phys. Rev. Lett.* 99 (2007), p. 217801.

[160] T. E. Ouldridge, A. A. Louis, and J. P. Doye. "DNA nanotweezers studied with a coarse-grained model of DNA". In: *Phys. Rev. Lett.* 104.17 (2010), p. 178101.

[161] T. E. Ouldridge, A. A. Louis, and J. P. Doye. "Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model". In: *J. Chem. Phys.* 134.8 (2011), 02B627.

[162] P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. K. Doye, and A. A. Louis. "Sequence-dependent thermodynamics of a coarse-grained DNA model". In: *J. Chem. Phys.* 137.13, 135101 (2012), p. 135101.

[163] L. Rovigatti, P. Šulc, I. Z. Reguly, and F. Romano. "A comparison between parallelization approaches in molecular dynamics simulations on GPUs". In: *J. Comput. Chem.* 36.1 (2015), pp. 1–8.

[164] B. E. Snodin, F. Randisi, M. Mosayebi, P. Šulc, J. S. Schreck, F. Romano, T. E. Ouldridge, R. Tsukanov, E. Nir, A. A. Louis, et al. "Introducing improved structural properties and salt dependence into a coarse-grained model of DNA". In: *J. Chem. Phys.* 142.23 (2015), p. 234901.

[165] S. W. Kowalczyk, D. B. Wells, A. Aksimentiev, and C. Dekker. "Slowing down DNA translocation through a nanopore in lithium chloride". In: *Nano Lett.* 12.2 (2012), pp. 1038–1044.

[166] F. Farahpour, A. Maleknejad, F. Varnik, and M. R. Ejtehadi. "Chain deformation in translocation phenomena". In: *Soft Matter* 9.9 (2013), pp. 2750–2759.

[167] A. Y. Grosberg and Y. Rabin. "DNA capture into a nanopore: interplay of diffusion and electrohydrodynamics". In: *J. Chem. Phys.* 133.16 (2010), 10B617.

[168] M Muthukumar. "Theory of capture rate in polymer translocation". In: *J. Chem. Phys.* 132.19 (2010), 05B605.

[169] M. Wanunu, W. Morrison, Y. Rabin, A. Y. Grosberg, and A. Meller. "Electrostatic focusing of unlabelled DNA into nanoscale pores using a salt gradient". In: *Nat. Nanotechnol.* 5.2 (2010), pp. 160–165.

[170] C. Micheletti, D. Marenduzzo, E. Orlandini, and D. W. Sumners. "Simulations of knotting in confined circular DNA". In: *Biophys J.* 95 (2008), pp. 3591–3599.

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

[172] G. Witz, G. Dietler, and A. Stasiak. "Tightening of DNA knots by supercoiling facilitates their unknotting by type II DNA topoisomerases". In: *Proc. Natl. Acad. Sci. U.S.A.* 108.9 (2011), pp. 3608–3611.

[173] J.-Y. Huang and P.-Y. Lai. "Crossings and writhe of flexible and ideal knots". In: *Phys. Rev. E* 63.2 (2001), p. 021506.

[174] K. Klenin and J. Langowski. "Computation of writhe in modeling of supercoiled DNA". In: *Biopolymers* 54.5 (2000), pp. 307–317.

[175] T. Ikonen, A. Bhattacharya, T. Ala-Nissila, and W. Sung. "Unifying model of driven polymer translocation". In: *Phys. Rev. E* 85 (5 2012), p. 051803.

[176] C. Plesa, N. van Loo, P. Ketterer, H. Dietz, and C. Dekker. "Velocity of DNA during translocation through a solid-state nanopore". In: *Nano Lett.* 15.1 (2014), pp. 732–737.

[177] T. Sakaue. "Nonequilibrium dynamics of polymer translocation and straightening". In: *Phys. Rev. E* 76 (2007), p. 021803.

[178] P. Rowghanian and A. Y. Grosberg. "Force-driven polymer translocation through a nanopore: an old problem revisited". In: *J. Phys. Chem. B* 115.48 (2011), pp. 14127–14135.

[179] J. Sarabadani, T. Ikonen, and T. Ala-Nissila. "Iso-flux tension propagation theory of driven polymer translocation: The role of initial configurations". In: *J. Chem. Phys.* 141.21 (2014), p. 214907.

[180] C. Forrey and M. Muthukumar. "Langevin dynamics simulations of ds-DNA translocation through synthetic nanopores". In: *J. Chem. Phys.* 127 (2007), p. 015102.

[181] T. Ikonen, A. Bhattacharya, T. Ala-Nissila, and W. Sung. "Influence of pore friction on the universal aspects of driven polymer translocation". In: *EPL* 103.3 (2013), p. 38001.

[182] S. Carson, J. Wilson, A. Aksimentiev, and M. Wanunu. "Smooth DNA transport through a narrowed pore geometry". In: *Biophys. J.* 107.10 (2014), pp. 2381–2393.

[183] L. Dai, C. B. Renner, and P. S. Doyle. "Metastable Tight Knots in Semiflexible Chains". In: *Macromolecules* 47.17 (2014), pp. 6135–6140.

[184] M. Caraglio, C. Micheletti, and E. Orlandini. "Stretching Response of Knotted and Unknotted Polymer Chains". In: *Phys. Rev. Lett.* 115.18 (2015), p. 188301.

[185] S. Kirmizialtin and D. E. Makarov. "Simulations of the untying of molecular friction knots between individual polymer strands." In: *J. Chem. Phys.* 128.9 (2008), p. 094901.

[186] V. Narsimhan, C. B. Renner, and P. S. Doyle. "Jamming of Knots along a Tensioned Chain". In: *ACS Macro Lett.* 5.1 (2016), pp. 123–127.

[187] D. Racko, F. Benedetti, J. Dorier, Y. Burnier, and A. Stasiak. "Generation of supercoils in nicked and gapped DNA drives DNA unknotting and postreplicative decatenation". In: *Nucleic Acids Res.* 43.15 (2015), p. 7229.