Polymer-theory insights into biomolecular systems

Salvatore ASSENZA

C'è una casetta piccola così...

Acknowledgements

i

Acknowledgements

Gracias al gran maestro de ceremonias Javi. Hombre, no sabría ni por donde empezar si tubiera que agradecerte todo lo que has hecho por mi. Me limitaré a darte las gracias por ayudarme poniendo lavadoras y por todo el apoyo que me has dado en los inevitables momentos malos durante estos cuatro años. Tu amistad es una de las cosas más preciadas que he ganado en este periodo (sin embargo, lo siento: no la más preciosa!).

Y gracias a Cristina, que me ha enseñado que hay cosas mucho más importantes que el trabajo. Ya sabes lo importante que eres para mi y habrá mejor lugar y momento para compartirlo con los demás. Aquí te agradezco sólo por haberme estado tan cerca durante este doctorado en general, y por haberme soportado mientras escribia mi tésis. Si veo el futuro con tanto optimismo es sólo porque estoy seguro que formarás parte de él.

Y por fin gracias a Paco, cuya guitarra me ha acompañado durante toda mi escritura de tésis.

On the train, somewhere between Lausanne and Zürich, April 1st, 2015   S. A.

Abstract

Boosted by the technological advances in experimental techniques, cellular biology is nowadays facing the need for quantitative approaches in order to rationalize the huge amount of collected data. A particularly succesfull theoretical framework is provided by Polymer Theory which, combined with molecular simulations, can capture the essential features of biomacromolecules and describe the cellular processes they participate to. This thesis provides a compendium of works showing the strength of this combination. In a first project, we model the twisting properties of amyloid fibrils by means of a simple coarse-grained approach, based on the competition between elasticity and electrostatic repulsion of nearby portions of the fibrils. The model quantitatively recapitulates the evolution of fibril periodicity as a function of the ionic strength of the solution and of the fibril width. A universal mesoscopic structural signature of the fibrils emerges from this picture, predicting a general, parameter-free law for the periodicity of the fibrils which is validated on several experimental results. A second work is focused on the role played by mitochondrial Hsp70 chaperone in the import of cytoplasmic proteins. Particularly, we computed by means of molecular simulations the effective free-energy profile for substrate translocation upon chaperone binding. We then used the resulting free energy to quantitatively characterize the kinetics of the import process and outline the essential role played by Hsp70 in this context. Finally, in a third project we studied the shape properties of a polymer under tension, a physical condition typically realized both in single-molecule experiments and *in vivo*. By means of analytical calculations and Monte Carlo simulations, we develop a theoretical framework which quantitatively describes these properties, highlighting the interplay between external force and chain size in determining the spatial distribution of a stretched chain.

Key words: Polymer Theory | Coarse-Grained Modelling | Amyloid Fibrils | Hsp70 Chaperones | Mitochondrial Import | Stretched Polymers

Sommario

Sotto la spinta dei progressi tecnologici nelle tecniche sperimentali, la biologia cellulare moderna si trova ad affrontare il bisogno di approcci quantitativi atti a descrivere l'immensa mole di dati raccolti. La fisica dei polimeri fornisce un contesto teorico particolarmente efficace e, in combinazione con simulazioni molecolari, è in grado di riprodurre le caratteristiche essenziali delle biomacromolecole e di descrivere i processi cellulari a cui prendono parte. In questa tesi riportiamo un compendio di lavori che mostrano la forza di questa combinazione. In un primo progetto, costruiamo un semplice modello coarse-grained sulle proprietà torsionali delle fibrille amiloidi, basato sulla competizione tra l'elasticità della fibrilla e la repulsione elettrostatica tra porzioni di quest'ultima vicine tra loro. Il modello riproduce quantitativamente l'evoluzione della periodicità della fibrilla in funzione della sua larghezza e della forza ionica della soluzione. Da questa competizione emergono delle proprietà universali a scale mesoscopiche, secondo cui le periodicità delle fibrille sono soggette a delle leggi generali non dipendenti da parametri fenomenologici. Tali leggi universali sono quindi confermate attraverso il confronto con diversi risultati sperimentali. In un secondo lavoro analizziamo il ruolo giocato dal chaperone Hsp70 nell'importazione di proteine del citoplasma all'interno dei mitocondri. Attraverso delle simulazioni molecolari calcoliamo l'energia libera efficace che regola la traslocazione di tali proteine in presenza del chaperone. Questo profilo di energia libera ci permette di caratterizzare a livello quantitativo le proprietà cinetiche del processo di importazione, enfatizzando il contributo fondamentale del chaperone. Come terzo e ultimo progetto studiamo le proprietà di un polimero sottoposto a una forza esterna, una condizione tipicamente realizzata in esperimenti di singola molecola e *in vivo*, focalizzandoci sulla sua forma tridimensionale. Attraverso calcoli analitici e simulazioni Monte Carlo riusciamo a delineare un contesto teorico che descrive quantitativamente tali caratteristiche e illustra in che modo la forza esterna e la taglia del polimero si bilanciano nel determinare la distribuzione spaziale di una catena sotto tensione.

Parole Chiave: Teoria dei Polimeri | Modelli Coarse-Grained | Fibrille Amiloidi | Chaperone Hsp70 | Importazione Mitocondriale | Tiraggio di Polimeri

Contents

Contents

List of Figures

List of Figures

List of Tables

Introduction

The last century has witnessed the literal outbreak of our knowledge in biological systems. Even restricting our focus to the subcellular realm, the rate of new achievements is so large that standard textbooks such as *Molecular Biology of the Cell* by Alberts *et al.* [1] need to be deeply revised every few years[1]. The wide range of different experimental techniques has led to the accumulation of a large amount of data. To cite just a few notable examples, NMR-retrieved structures nowadays provide information on the three-dimensional arrangement of crystallized proteins with a precision on the order of Angstroms [2]; the Hi-C technique gives access to a genome-wide contact map of eukaryotic chromatin [3]; the employment of green-fluorescent proteins [4] permits to dynamically monitor several aspects related to protein production, like *e.g.* genes activity bursts [5]; single-molecule techniques such as FRET [6, 7], AFM [8, 9] or Optical Tweezers [10] have greatly enhanced our knowledge of conformational and elastic properties of biological polymers.

This gargantuan production of data has led modern cellular biology to the possibility of and the need for quantitative modelling approaches, as explicitely stated by Alberts and coworkers [1]:

> The information we have [...] is largely qualitative rather than quantitative. Most often, cell biologists studying the cell's control systems sum up their knowledge in simple schematic diagrams rather than in numbers, graphs, and differential equations. To progress from qualitative descriptions and intuitive reasoning to quantitative descriptions and mathematical deduction is one of the biggest challenges for contemporary cell biology.

This call of duty has been responded to by many scientists from several fields. Attracted by the increasing abundance of data and the wide variety of mechanisms and structures proposed by biologists, they dedicated their research to investigating biosystems with the approaches typical of their own disciplines. Particularly, both theoretical and experimental physicists have strongly become interested in bio-related problems. From the experimental point of view, physicists make large use of single-molecule techniques. The latter, as opposed to traditional biochemical approaches, have the unique feature that they allow the researcher to focus on individual agents acting in the cellular environment, which is more akin to how a

---

[1]This book is now at its sixth edition and was first published in 1983.

physicist usually works (for an introduction to single-molecule experiments, see *e.g.* [11] and references therein). As a result, there are at present many *in vitro* studies which characterize the conformational and mechanical properties of macromolecules. For example, the elastic response of double-stranded DNA to an external force has been probed by means of Magnetic Tweezers [12]; mechanical unfolding of single domains of multimodular proteins has been analyzed by AFM-pulling measurements [13]; misfolding of multidomain proteins has been directly observed in FRET experiments [14].

From the theoretical point of view, the typical energies characterizing biological processes [11] (order of $k_B T \sim 10^{-21}$ J) and their length scales (order of nanometers, thus larger than the onset of quantum effects [15]) identify Classical Statistical Mechanics as the most suitable framework to study biological macromolecules. In this regard, useful insight can be gained from Polymer Theory [16]. The latter starts from the basic assumption that most of the macroscopic[2] properties of a polymer are mainly due to its wire-like structure, while the microscopic details serve only to set the physical parameters - such as the linear charge density, bending and twisting rigidities, etc. - describing it in the coarse-grained picture. Since both proteins and nucleic acids are chains constructed by covalently binding some basic bricks (amino acids and nucleobases respectively), they are expected to follow the general laws predicted by Polymer Theory. As several works have shown, this is indeed the case. For example, the force-vs-extension measurements on double-stranded DNA [12] are quantitatively described by means of the so-called Wormlike Chain, a simple model from Polymer Theory which depicts DNA as a continuous wire with a given bending rigidity [17]. In another work [18], it has been shown that the Stokes' radius of denatured and intrinsically disordered proteins closely follows the scaling relationship with their contour length predicted by Polymer Theory. This result thus rationalizes the unfolding process as a coil-to-globule phase transition, where according to the nature of the solvent the polymer is best described as a swollen coil or a compact sphere [15]. It is worth mentioning that the benefits go also the other way round. Indeed, while it is true that Biology has profited from the appearance of biophysicists, it is also boosting further research in fundamental questions of Statistical Mechanics. Particularly, biological macromolecules are suitable systems to benchmark recent theories on the out-of-equilibrium thermodynamics [11].

Another important toolset to address the need for quantitative approaches in Biology is today available thanks to the growth of computational power and to the development of simulation algorithms. Molecular Dynamics (MD) and Monte Carlo (MC) simulations are nowadays widely-used tools to explore the behavior of living matter at different levels of resolution. Just to get a feeling for numbers, in [19] the authors observe that, in a time window covering the first six months of 2002, the popular database ISI Web of Science [20] returned more than 250 results under a query with keywords "Molecular Dynamics" and "protein". A similar search conducted in December 2014 returns more than 2,000 results[3]. To cite a few examples, all-atom MD

---

[2]In this context, "macroscopic" refers to global features of a polymer, while "microscopic" focuses onto atomic-scale properties.

[3]In contrast, "only" ~150 papers were found by using the keywords "Monte Carlo" and "protein".

simulations are successfully used to explore specific microscopic mechanisms not accessible to experiments, such as the behavior of external atoms within folded proteins (*e.g.* oxygen in myoglobin [19]) or particle transport through membrane pores [21]. Moreover, they are also widely employed to refine the structures retrieved from NMR experiments [22,23]. Monte Carlo simulations are used as well in a large variety of flavors, including *e.g.* sequence-alignment algorithms [24], phylogeny inference [25] and simulation of subcellular processes [26].

Within Polymer Theory, MD and MC simulations provide complementary tools to explore both conformational and dynamical coarse-grained properties of biological macromolecules. Particularly, while MC samplings are tipically faster in encompassing the relevant regions of the conformational space, MD simulations have the advantage of being usually more easily adapted to increasing levels of complexity. Thus, if one wants to study the conformational properties of simple polymer models such as the Freely-Jointed Chain (FJC), where the macromolecule is described as a chain of segments of fixed length with random orientations [27], the best strategy is to use a suitable Monte Carlo approach. In contrast, if one seeks to explore the phase space of chains described by more complicated Hamiltonians (like *e.g.* force-fields developed for specific classes of polymers) or subjected to complicated constraints, then it might be wiser to opt for MD simulations.

This Thesis provides an example of how theory and simulations can be fruitfully combined to describe biological macromolecules. Though focusing on quite diverse systems, the take-home message pervading the whole Thesis is that a coarse-grained description of a biopolymer can give very useful insights, which in some instances can also quantitatively account for experimental data. In Chapter 1, we present a simple model for the twisting properties of amyloid fibrils. In this case, Polymer Theory is mostly used to justify the assumptions of the model, which is based on the minimization of the total energy per fibril. The model predicts a general, parameter-free law for the periodicity of the fibrils together with an explicit formula based on the interplay between elastic and electrostatic contributions, which are both in very good quantitative agreement with experimental results. In Chapter 2, we focus on the chaperone-driven import of proteins into organelles. By means of coarse-grained MD simulations, we compute the effective free-energy profile for the translocation process due to the presence of Hsp70 chaperones. The resulting free energy is then used to quantitatively assess the kinetics of the import process, highlighting the essential role played by Hsp70 in importing cytoplasmic proteins. In Chapter 3, we focus on the shape of a stretched chain. Though not being focused on a specific biomolecular system, our results are of biological interest since biopolymers are subjected to mechanical stress in many cellular processes, and their shape affects the way they interact with other molecules. Characterizing a stretched chain by means of analytical calculations and MC simulations, we are able to develop a theoretical framework which quantitatively captures all the essential features of both the shape and the orientation of the polymer within the simple case of a FJC.

1 Universal behavior in the mesoscale properties of amyloid fibrils

1.1 Amyloid fibrils: the general picture

1.1.1 Protein folding depends on external conditions

Within the cellular microuniverse, proteins are the factotums, being involved in "virtually every process taking place in a living organism" [28]. Their employment embrace the most disparate functions, ranging from membrane structural support to targeting of specific molecules, as well as molecular transport and signal transmission [1, 28]. The importance of proteins is reflected in their relative abundance in the cellular environment as compared to other molecules. Indeed, more than a half of the dry weight of a cell is typically due to polypeptides [29] (a census of the relative quantities of various molecules in the bacterium *E. Coli* is reported in Table 1.1).

A fundamental role in the correct functioning of a protein is played by its three-dimensional shape [1], which emerges from the complex pattern of residue-residue and residue-solvent interactions and is thus ultimately determined by the amino-acids sequence and the solvent

Molecule	% of dry weight	# of molecules
Protein	55.0	2.4×10^6
RNA	20.4	2.6×10^5
Phospholipid	9.1	22×10^6
Small Molecules	3.9	\
Lypopolysaccharide	3.4	1.2×10^6
DNA	3.1	2
Murein	2.5	1
Glycogen	2.5	4.4×10^3

Table 1.1 – Weight distribution and absolute quantities of the various molecules observed in an *E. Coli* cell (data from [29]).

conditions. Roughly speaking, the final shape of a folded protein has a bulk formed by hydrophobic residues and the solvent-exposed surface mainly made of polar or charged amino acids [28]. This structure is stabilized by the presence of residue-residue interactions such as H-bonds, the latter being responsible of the formation of typical local (*e.g. α*-helices) and nonlocal (*e.g. β*-sheets) secondary structure.

From the point of view of Polymer Theory, a folded protein is entropically inconvenient, so that the interactions mentioned above have to sum up to a strong enthalpic term U in order to counterbalance the entropy loss. The relative strength of the two competing terms in the free energy F depends on temperature T by means of the well-known formula $F = U - TS$, where S is the entropy of the system. For low values of T, the interaction energy dominates over the entropy, while for high temperatures S provides the main contribution to the free energy. As a result, it comes as no surprise that by increasing the temperature a folded protein can swell [28]. In Polymer-Theory jargon, in the low-temperature regime water behaves as a *poor solvent* [27] and as a result the protein folds into a compact structure. In contrast, in the high-temperature, entropy-dominated regime water behaves as a *good solvent*, and the typical conformation of the protein is that of a random coil [27].

Since we are interested in biological environments, we normally focus on the so-called *physiological conditions, i.e.* the typical solvent status found within a cell. Namely, for our purposes such conditions consist in considering water at room temperature (typically in the range $20 - 40$ °C) and at roughly neutral pH ($6 - 8$) [1]. It does then make sense to speak of *native structure* of a protein, indicating with this term the minimum free-energy state at physiological conditions or, in other words, the "working" conformation. According to whether in physiological conditions water is a poor or good solvent for a given protein, its native structure is folded or intrinsically disordered, respectively.

1.1.2 Protein destabilization can lead to aggregation

The three-dimensional features of a protein depend on its sequence as well as on solvent conditions. The structure of a natively-folded protein can be destabilized by raising temperature, as we mentioned in the previous section, or by adding chemical denaturants such as guanidinium chloride or urea [30, 31]. Moreover, *in vivo* proteins live in a very crowded environment, so that interactions with other molecules can contribute to their destabilization as well [29].

Usually, in physiological conditions proteins are quite resistant to destabilization, and one needs to put them through artificially-harsh conditions in order to cause their swelling. Nevertheless, in some instances the destabilization takes place even at physiological conditions [32]. In most cases, the affected proteins remain inert and are eventually eliminated by specific cellular processes [28]. However, in some instances the destabilized proteins do interact with each other or with other molecules, though not in the way they were intially conceived for. Particularly, when many corrupted proteins are present in a system, they can undergo an

uncontrolled aggregation process known as *fibrillation* [32]. Though the precise aggregation pathways of many fibrillar systems are still under debate, one can generally say that this process takes place when a destabilized protein shows on its surface a hydrophobic region, which makes it prone to aggregation in a nucleation-like fashion [32]. The resulting products, known as *amyloid fibrils*, have recently attracted a lot of attention in the clinical literature, due to their involvement in neurodegenerative disorders such as Alzheimer's, Parkinson's or Huntington's diseases, together with the so-called prion-related diseases [32–36].

1.1.3   Amyloid fibrils are ubiquitous multiscale objects

Remarkably, there are many proteins which can serve as starting monomers for the fibrillation process, and they do not share particular properties. They can be natively unfolded polypeptides or globular proteins. Their sequences have not significant statistical overlaps [33] and have lengths spanning from tens to hundreds of residues [32]. Moreover, also their secondary structure is strongly diversified [33].

In spite of the heterogeneity of the starting monomers, all fibrils share an invariant structural fingerprint at the Angstrom length scale. Particularly, independently of the specific peptidic primary sequence composing them, the final fibril is very rich in β-sheets structure, displaying a core formed by longitudinal sheets made of cross-β-strands [33, 37]. Moreover, all fibrils are rope-like structures, where several protofilaments twist around each other. Still, provided suitable solvent conditions, sometimes very far from the physiological ones [38–44], amyloid fibrils have been formed *in vitro* from disease-associated as well as from disease-unrelated proteins [45, 46]. This ubiquity, together with their relevance in biotechnology and nanotechnology applications [47–49] and their uncommon set of physical and structural properties [50–53], attracted the attention of scientist coming from disciplines as diverse as structural biology, chemistry, physics and materials science.

While at the Angstrom length scale amyloid fibrils show undeniable universal properties (that is, the β-sheets pattern mentioned above), at larger length scales their forms of polymorphism diverge [54–57], as does the view on the main physical mechanisms ruling amyloid mesoscopic structure [58–65]. It is the structural architecture at these length scales that controls some of the ensemble physical properties, ranging from their persistence length to their liquid-crystalline properties in three and two dimensions [66]. These features are epitomized by the existence of a very regular structure in amyloid fibrils at a length scale in the order of tens to hundreds of nanometers [43, 54–57].

For example, Adamcik et al. [43] have performed a detailed statistical analysis of Atomic Force Microscopy (AFM) images of fibrils produced *in vitro* starting from β-lactoglobulin (βlg) monomers, finding that they are organized as twisted ribbons formed by laterally attaching single protofilaments, and that different populations of fibrils can be identified depending on the number of protofilaments composing them.

In the present chapter, we introduce an original model [67] where the twisting structure emerges as a compromise between opposing forces. As we will show, our approach not only allows a faithful mapping of the amyloids structural features, but also highlights, for the first time, a common universal behavior in their mesoscopic topological signature.

Before delving into the derivation of our model and its comparison with experimental data, we will spend the next two sections in reviewing in some detail the findings in [43], where the geometry of βlg fibrils was first unveiled, and the results in [44], where the key role played by electrostatic interactions in determining the quantitative values of their twisting properties was highlighted.

1.2 βlg fibrils are stiff twisting ribbons

The packing arrangement of βlg fibrils was studied in [43] at the single-molecule level by means of AFM measurements. From a statistical analysis of the height fluctuations of fibrils deposited on mica, the authors show that these aggregates are composed by laterally-associated protofilaments with the packing arrangement of a flat ribbon, and propose a simple mechanism responsible for the twisting structure.

1.2.1 Fibrils Preparation

The preparation of fibrils started from purified βlg powder, provided by Davisco Foods International. The steps followed to obtain the fibrils were the following:

1. Dissolve the powder in Milli-Q water at room temperature in order to prepare a 2 wt%[1] solution.

2. Centrifuge the solution for one hour at 20 °C and $10800g$.

3. Adjust the solution at $pH = 2$.

4. Put the solution in flasks and heat at 90 °C for five hours, stirring it during heating.

5. Cool the flasks immerging them in ice-water mixtures in order to quench the aggregation process.

6. Dilute the βlg fibrils solution at 0.1 wt% by using $pH = 2$ Milli-Q water.

The destabilization of a natively-folded βlg monomer is in this case achieved thanks to the increase in temperature performed in step 4. As a final result, a 0.1 wt% acid solution of fibrils was obtained.

---

[1]wt% indicates the percentage of solute mass with respect the total weight of the solution.

Afterwards, 10 μl of this solution were deposited onto freshly cleaved mica, incubated for two minutes, rinsed with Milli-Q water and dried by nitrogen. In order to improve resolution, the images used for the statistical analysis were taken on dried samples. However, in order to rule out possible artifacts due to the drying process, a countercheck was done by performing complementary experiments maintaining the wet environment.

1.2.2 Height distribution points towards a ribbon-like arrangement

After having deposited the βlg fibrils onto the substrate, the latter was scanned by means of an AFM device working in tapping mode (for an introduction to AFM see [8]). In fig.1.1 we report typical AFM images at different levels of magnification. The fibrils have contour lengths on the order of microns. When compared to the typical size of a βlg monomer (around 2 nm), these values stress the huge extent of the aggregation process. In contrast, the typical width falls in the nanometer scale, thus outlining a strongly-elongated shape for the fibrils. Nevertheless, due to the poor lateral resolution of AFM [8], all the subsequent analysis is based on height measurements. In fig.1.2 we report the frequency distributions of fibrils contour lengths (a) and heights (b). As anticipated, the typical length of a fibril is in the micron length scale, ranging from 0.5 to 15 μm, with a single maximum around 5.5 μm. In contrast, the heights follow a multimodal distribution, showing peaks at the quantized values 2, 4, 6, 8 and 10 nm (the latter two are not easily distinguishable because of the lack of statistical populations). It is worth noting that the height along each fibril shows fluctuations (corresponding to periodic intensity changes in the AFM pictures in fig.1.3) which are localized in regions of the former distributions corresponding to specific peaks, with the maximum heights recorded in correspondence to the peaks. Furthermore, within each fibril these fluctuations were systematically observed as long as the fibrils under inspection had maximum height above 3 nm.

All the previous observations strongly point towards a twisting structure of fibrils, and are compatible with a flat-ribbon packing arrangement, where n protofilaments are laterally associated (see fig.1.4). In this picture, the multimodal heights distribution emerges from the presence of several fibrils populations corresponding to different values of n. The lowest observed maximum height (2 nm) is relative to the case $n = 1$, for which no periodicity is observed. Instead, the larger values are relative to fibrils with $n > 1$ and are multiples of 2 nm. Therefore, the maxima corresponding to 4, 6, 8 and 10 nm highlight the presence of fibrils made of $n =$2, 3, 4 and 5 protofilaments. Moreover, also the localization of the height maxima on the peaks of the distributions can be rationalized into a ribbon-like arrangement. By extending the analysis presented in [43], let θ be the angle formed by the substrate and the section of a fibril made of n protofilaments. Since the fibril is a twisting object, the value of θ depends on the position considered along the fibril itself. Particularly, assuming a homogeneous twist the angle θ is expected to be uniformly distributed in the range $[0, \pi/2]$. Therefore, imposing normalization gives for its probability distribution $P(\theta) = 2/\pi$. If $b = 2$ nm is the diameter of a protofilament, the fibril height h corresponding to a given value of θ can be easily computed
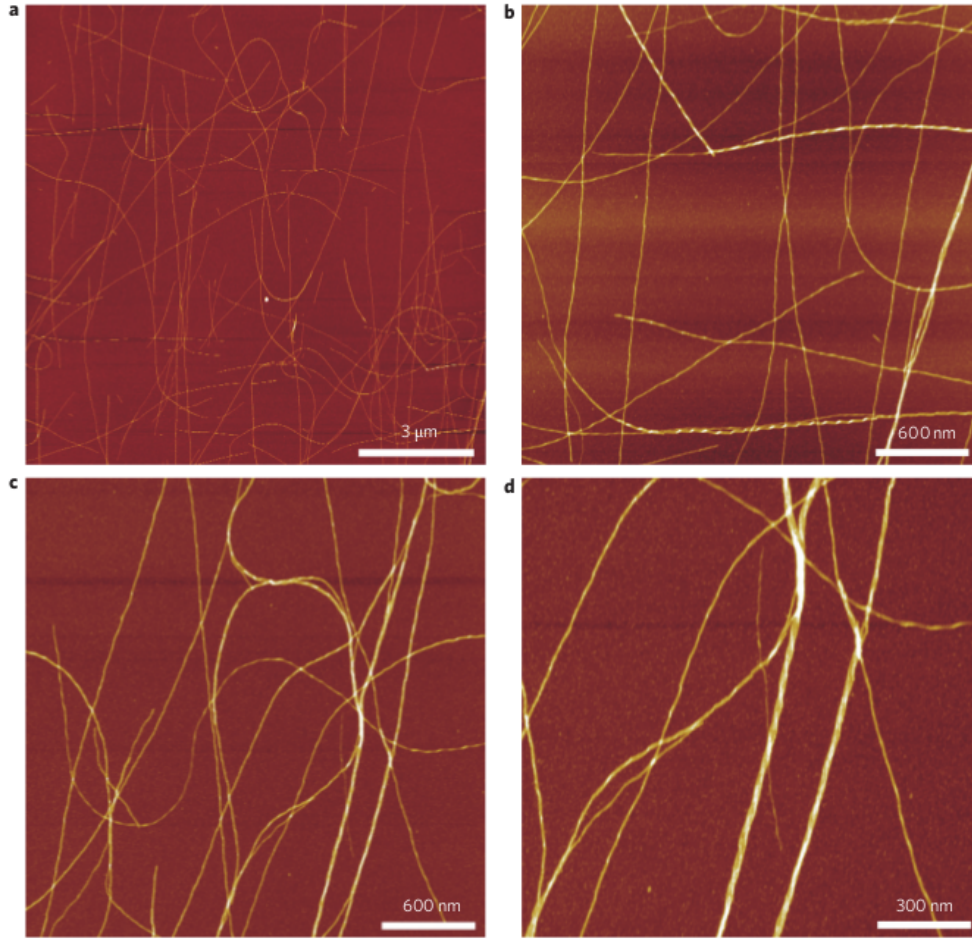
Figure 1.1 – AFM images of βlg fibrils on mica dried from a pH=2 solution at different levels of magnification. Reprinted from [43].

as $h = b + (n-1)b\sin\theta$. By defining $x = h - b$, which corresponds to the difference between the height of the fibril at the point under inspection and its minimum value (*i.e.* when the fibril is adsorbed horizontally), we find $x = (n-1)b\sin\theta$. Inverting the previous formula gives

$$\theta = \arcsin\left[\frac{x}{(n-1)b}\right]. \tag{1.1}$$

By operating a change of variable $\theta \to x$, we finally find for the distribution $\bar{P}(x)$ of the height difference

$$\bar{P}(x)dx = P(\theta)d\theta = P(\theta)\frac{d\theta}{dx}dx = \frac{2}{\pi}\frac{1}{(n-1)b\sqrt{1 - \left[\frac{x}{(n-1)b}\right]^2}}dx. \tag{1.2}$$

The previous formula shows that a homogeneous twist results in a monotonically increas-
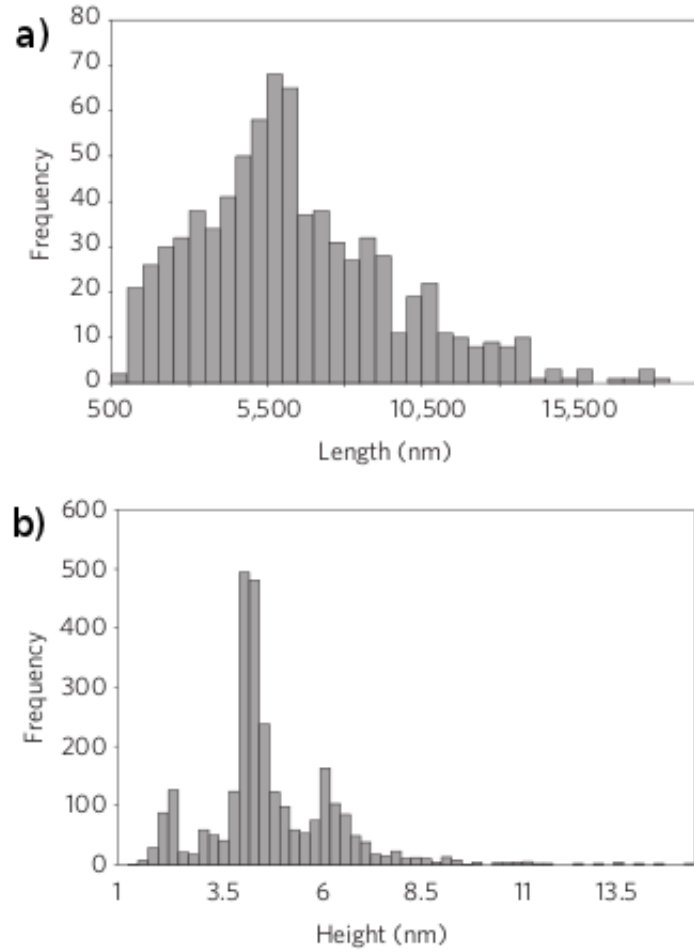
Figure 1.2 – Contour length (a) and fibril height (b) statistical analysis. Adapted from [43].

ing distribution of heights. Intuitively, when the ribbon is closest to its maximum-height configuration its variation in height slows down, so that a uniform sampling of the height along the fibril contour length results into a richer population of almost-vertically-adsorbed configurations. In fig.1.5a we plot eqn (1.2), while in fig.1.5b an example of the corresponding histogram obtained from the experimental data is reported, restricted to the population centered around 6 nm in fig.1.2b. The qualitative agreement between the predicted distribution and the experimental histogram gives a further argument in favor of a flat ribbon as the correct packing arrangement of βlg fibrils.

1.2.3   Analysis of data: halfpitch and persistence length

In order to describe the geometric and mechanical properties of βlg fibrils, there are three relevant length scales to consider: the diameter of a protofilament $b = 2$ nm, the period of
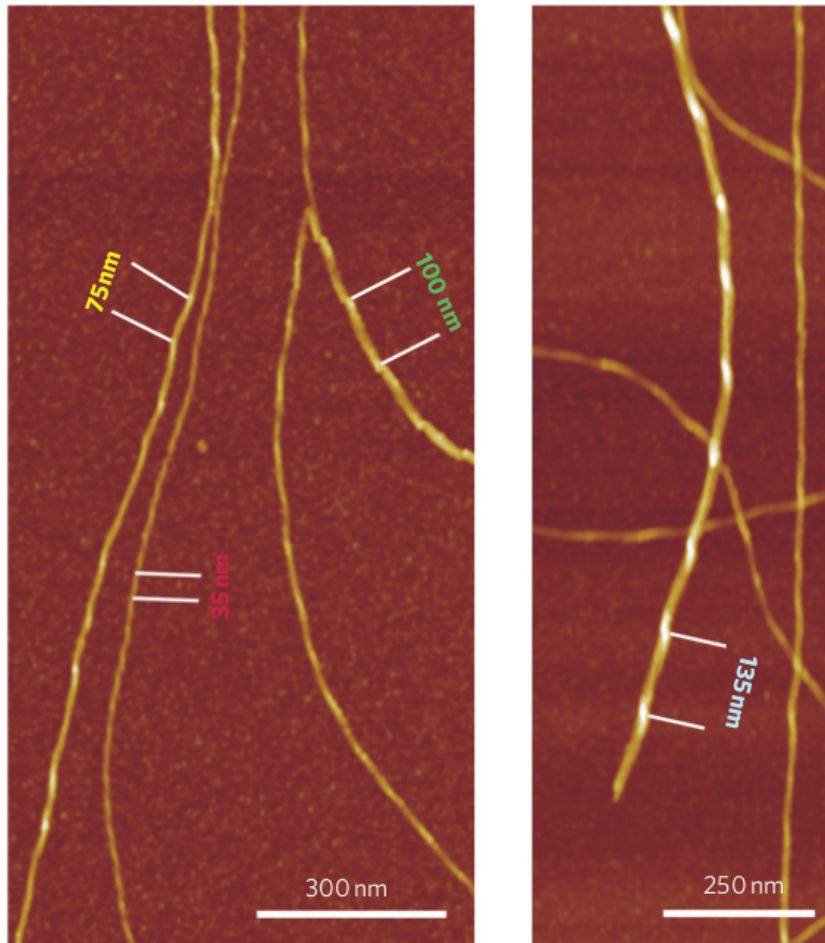
Figure 1.3 – Fibrils thicker than 3 nm show periodic height fluctuations. The thicker the fibril, the larger the period. Reprinted from [43].

height fluctuations and the persistence length of the fibril as a polymer. While the first length scale was already introduced in the previous section, we now focus on the remaining two.

The period of the height fluctuations can be measured as the distance between two consecutive maxima and corresponds to the total length L travelled by the fibril before having twisted for half a turn (total twist angle equal to π). We will refer to this quantity as *halfpitch*. The average values corresponding to the different populations are reported on representative AFM images in fig.1.3. In fig.1.6 these values are plotted as a function of the maximum height of the fibrils, displaying an approximately linear behavior.

Again, the previous results can be rationalized by considering a ribbon-like packing arrangement. In this regard, the authors in [43] proposed that the twisting of the fibrils emerges as a compromise between their elasticity and the electrostatic repulsion between nearby portions along the fibrils, the latter due to the large number of charges present on their surface [68]. Let
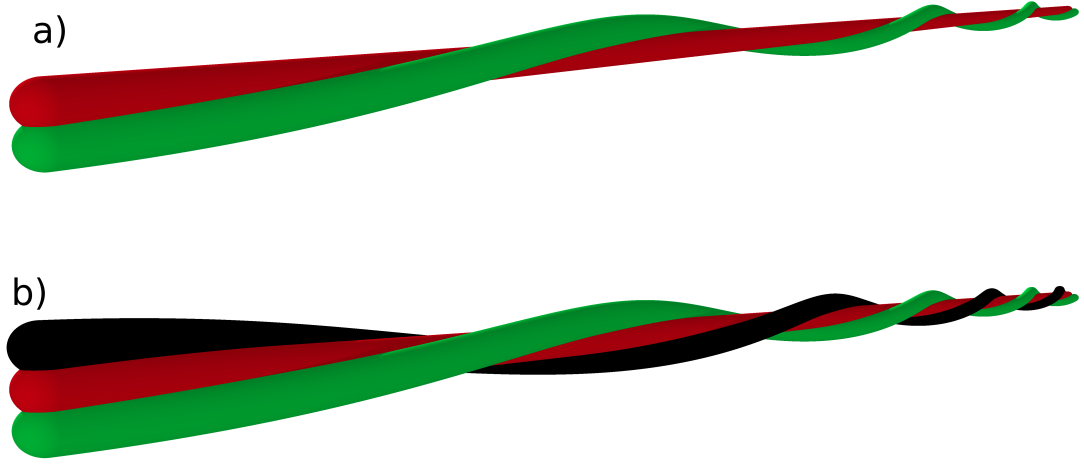
Figure 1.4 – Examples of twisting-flat-ribbons packing arrangements for $n = 2$ (a) and $n = 3$ (b).

us consider on a given fibril two transverse sections very close to each other. Because of the twist, these sections are tilted by a certain angle ϕ. From the elastic point of view, the optimum tilt angle would be $\phi = 0$, corresponding to a relaxed ribbon. In contrast, the electrostatic repulsion between the two sections would be minimized for $\phi = \pi/2$. Therefore, an intermediate equilibrium value of ϕ is expected. As a further ansatz, assume that the sections tilt until their ends reach a distance d, independent of the width of the fibril. This distance can be thought of as the threshold above which the structure of the fibril is disrupted. Assuming ϕ to be small, for a fibril made of n protofilaments simple geometric arguments give $d = (n-1)b\phi/2$. By inverting the previous formula we thus find

$$\phi = \frac{2d}{(n-1)b} \tag{1.3}$$

If in a halfpitch there are m transverse sections, the relation $m\phi = \pi$ must hold. Substituting (1.3) in the previous formula, we finally find

$$m = \frac{\pi b}{2d}(n-1) \tag{1.4}$$

Within this simple scheme, the halfpitch L of the fibril is proportional to m, so that the proportion $L \propto n$ holds, in agreement with the results reported in fig.1.6. As we will see later, a more refined model based on the same assumptions predicts that this relation is not exactly linear, though the resulting functional form is in practice very close to it.

The third fundamental length scale in the system is the persistence length of the fibrils. This quantity is a basic concept typically used in Polymer Theory to characterize the mechanical properties of polymers. Roughly speaking, the persistence length l_p provides the distance one
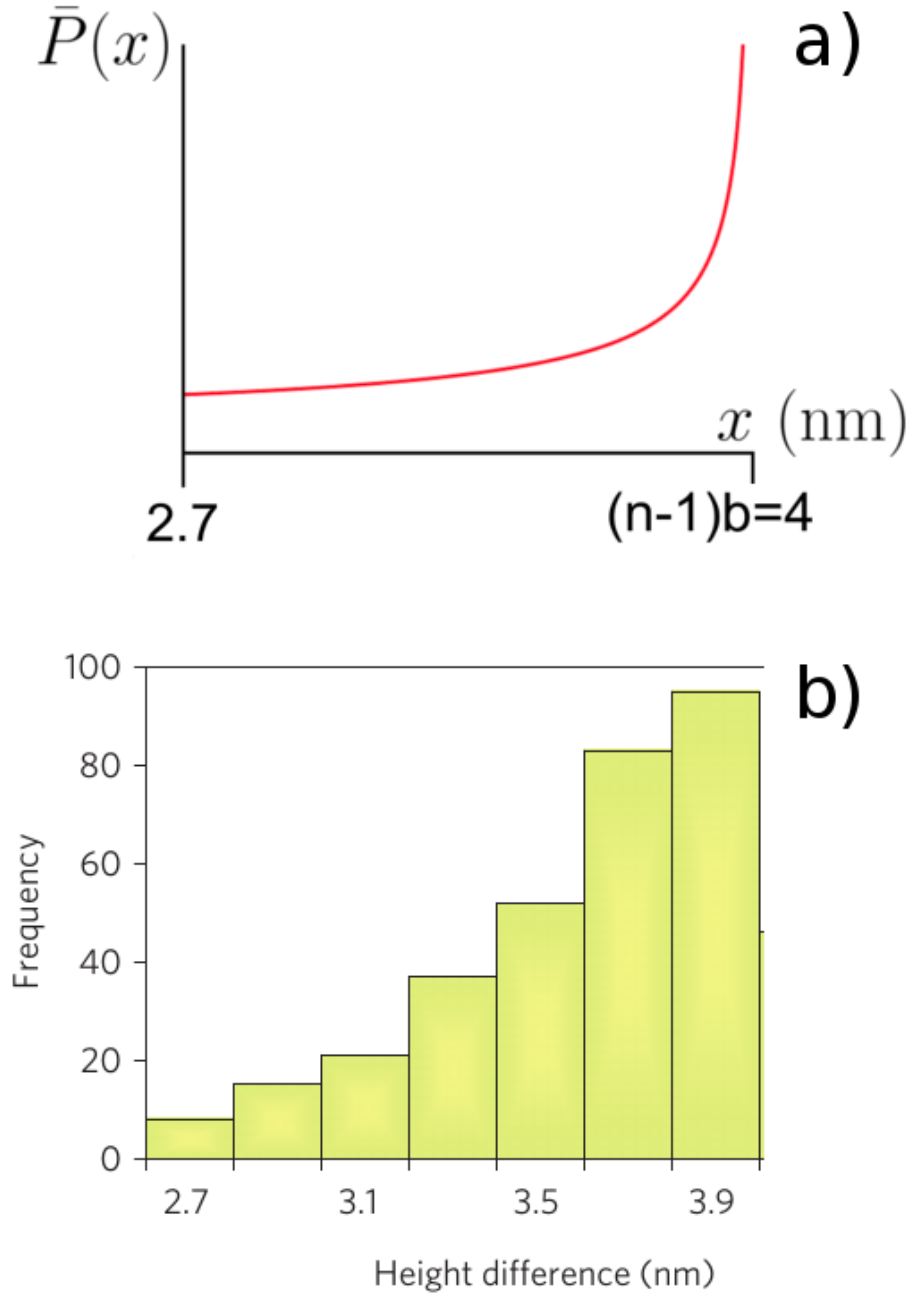
Figure 1.5 – Theoretical probability density (a) and experimental histogram (b, adapted from [43]) for the difference between the local and minimum height of a given fibril. Data in b) are restricted to the population relative to the peak at 6 *nm* in fig.1.2b.

has to travel along the polymer to observe appreciable bending due to thermal fluctuations. In other words, at length scales much lower than l_p the polymer behaves like a rigid rod, while at very large scales a random coil conformation is observed [27]. Having a look at fig.1.1 and
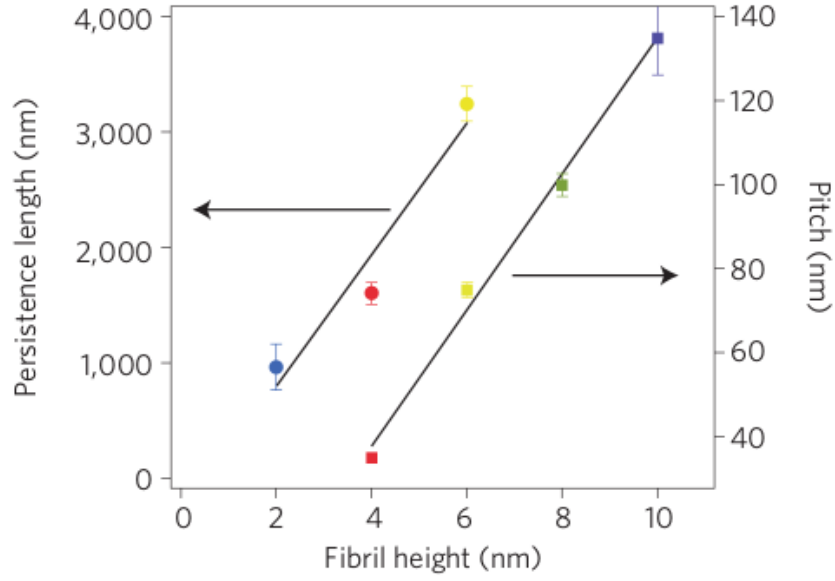
Figure 1.6 – Experimental values of persistence length (circles) and halfpitch (squares) versus fibril height. Although the number of points is quite restricted, both L and l_p seem to be well described by a straight line. Each value on the abscissa axis corresponds to a peak in fig.1.2. Adapted from [43].

fig.1.3, we can expect values of l_p ranging between hundreds to thousands of nanometers.

For a more precise definition, let us represent a given fibril as a smooth curve embedded in a $\mathscr{D}$-dimensional space (see the top panel in fig.1.7). For each arc length s, let $\theta(s)$ be the angle formed by the tangent vectors of the polymer at two points separated by a contour length s. By making a s-large window slide along the molecule, one obtains different values of $\theta(s)$. According to Polymer Theory [27], the average value of the cosine of this angle $\langle \cos\theta(s) \rangle$ (that is, the scalar product between the considered tangent vectors) defines a *correlation function*, which usually shows an exponential decay due to the Markovian character of the polymer [15]. By definition, the *persistence length l_p* is the length scale of this exponential decay: $\langle \cos\theta(s) \rangle = \exp(-s/l_p)$. Since it is related to the transverse fluctuations of a polymer, the persistence length is expected to depend on the dimensionality of the system [69]. It is customary to refer to its three-dimensional value as *the* persistence length of the polymer, and to suitably correct the exponential decay according to the dimensionality of the system. For example, in the present case the fibrils have been adsorbed onto a substrate, thus the embedding space is two dimensional and a factor of 2 needs to be introduced in the correlation function [69]:

$$\langle \cos\theta(s) \rangle = e^{-\frac{s}{2l_p}} . \tag{1.5}$$

15
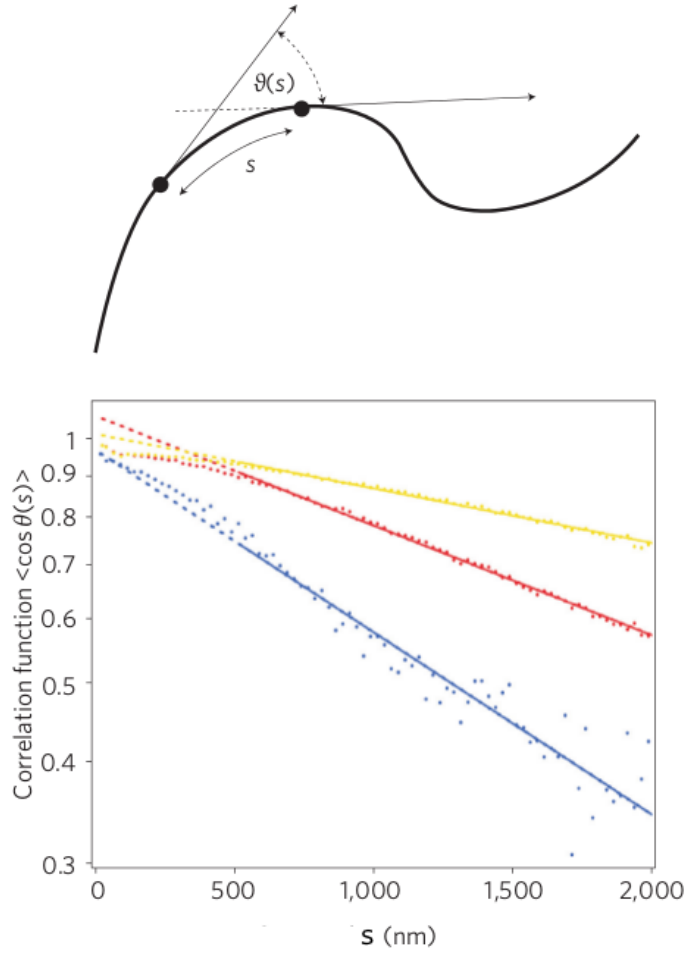
Figure 1.7 – Top: Sketch of a semiflexible polymer where we show the arc length s and the angle $\theta(s)$ between the tangents at points with contour distance s, used to define the correlation function. Bottom: experimental data of the average values of the correlation function versus the arc length s, and relative fits according to eqn (1.5). Blue, red and yellow points and lines correspond to fibrils with average height $2, 4$ and 6 nm respectively. Adapted from [43].

In this way, by computing the average of the measured values of the angle, one can extract the persistence length of the fibrils. Unfortunately, fibrils corresponding to $n = 4, 5$ are not sufficiently populated to provide reliable estimates of their l_p. In the bottom part of fig.1.7 the final average cosines are reported in log-lin scale for different values of the arc length s. The blue, red and yellow points correspond to fibrils belonging to populations centered at $2, 4$ and 6 nm respectively. As expected, the three classes of points are well fitted by an exponential decay, which in log scale becomes a straight line. From the slope of this line and making use of eqn (1.5), it is finally possible to extract the values of l_p, which are reported in fig.1.6. From the theoretical point of view, some scaling arguments point towards a linear relationship between l_p and n. Indeed, assuming the fibril to be a homogeneous beam with Young's modulus E, the

persistence length can be written as [27]

$$l_p = \frac{EM}{k_B T},$$ (1.6)

where k_B the Boltzmann constant, T the temperature and M the area moment of inertia of the section of the fibril (see below). Note that l_p is proportional to the Young's modulus of the polymer, which quantifies its intrinsic stiffness, and inversely related to temperature, since for hotter baths the thermal fluctuations overcome more easily the bending rigidity of the fibril. By modeling the section of the ribbon as a rectangle with sides b and nb and considering only the lowest mode to be involved in the thermal bending, the area moment of inertia can be easily computed as [70]

$$M = \int_{-\frac{nb}{2}}^{\frac{nb}{2}} dx \int_{-\frac{b}{2}}^{\frac{b}{2}} dy\, y^2 = \frac{1}{12} nb^4$$ (1.7)

Substituting (1.7) in eqn (1.6) we find

$$l_p = \frac{Eb^4}{12 k_B T} n \propto n,$$ (1.8)

in agreement with the linear behavior of the experimental data (fig.1.6).

All the arguments reported thus far provide compelling evidence that the structure of βlg fibrils is that of a twisting ribbon. Note that, by comparing the maximum and minimum heights, it is possible to extract the number of strands forming the fibril under investigation.

1.2.4 The electrostatic properties of the fibrils determine the value of L

The mechanism proposed in [43], according to which the twisting structure of fibrils arises as an interplay between elasticity and electrostatics, was experimentally tested in a subsequent work [44]. The basic idea was to measure the halfpitches of fibrils in different salinity conditions, by means of the same analysis performed in [43] and reported in the previous section. Indeed, adding salt into a solution increases the concentration of free ions, and as a result the net electrostatic interaction between nearby sections of the fibril is weakened [71]. Therefore, if electrostatics is responsible for the twisting structure of the fibril, its screening should lead to more relaxed ribbons. As a result, the halfpitch is expected to be an increasing function of the concentration of salt added to the solution. In practice this was achieved by forming fibrils with the same protocol as in [43] and adding salt at a later stage. This detail in the protocol is of great importance, since the electrostatic interaction plays an important role in fibril formation [55, 72, 73] and starting the aggregation process in different salinity conditions would ultimately result into structurally-different fibrils. In contrast, adding the salt only after the fibrillation has taken place affects only the twisting properties.

In [44], the authors added several concentrations of salt into solutions with preformed βlg

fibrils and let the system equilibrate for three weeks. In fig.1.8 we report the values of the halfpitches obtained for the various fibril populations adsorbed on mica at different values of salt concentration. Five different solution conditions were considered, where $0, 25, 50, 75$ and 100 mM of salt were added[2]. Considering that the experiments were carried at pH=2 (obtained by dissolving 10 mM of HCl into the solution), the total concentration I of free ions, known as *ionic strength*, was equal to $10, 35, 60, 85$ and 110 mM respectively.

In fig.1.8, panels belonging to the same column represent different fibris populations (*i.e.* different n) for given salinity conditions. Particularly, the salt-free case reproduces the same experimental conditions as in [43], and consequently the same values of the halfpitches are retrieved. By focusing on the rows of the mosaic, we can explore the behavior of the halfpitch for a given fibril population as a function of the ionic strength. As anticipated, L increases monotonically with I, thus conclusively confirming that electrostatic effects affect the twisting properties of the fibrils in the way qualitatively predicted in [43].

Finally, we note that for large ionic strength the fibrils appear more flattened. The periodic behavior of height fluctuations is still present, but it is not as smooth as for lower values of I. As a result, for wider ribbons a vertical adsorbed configuration is not anymore sustainable, and the fibril ends up being practically smashed on the surface.

1.3 A coarse-grained model for the twisting properties of amyloid fibrils

As we have seen above, the proposed mechanism responsible for the twisting of fibrils is based on the competing interplay between elasticity and electrostatic repulsion [43]. From an experimental perspective, the key role played by electrostatics was outlined by measuring the halfpitch of several populations of fibrils at different concentrations of added salt, which confirmed that stronger electrostatic interactions lead to more twisted structures [44]. In [44], the authors also propose a theoretical framework to rationalize the data, which captures the general trend of the halfpitches as a function of salinity. However, the quantitative agreement with data is not fully satisfactorily, and the dependence on fibrils populations has to be introduced by adapting a fitting parameter, which takes different values for different populations [44]. Nonetheless, this work paved the way for a more accurate approach, ultimately leading to our original model introduced and discussed in the present section [67].

1.3.1 Building up the model

As shown in fig.1.9, each protofilament in the ribbon (represented in the figure as a continuous strand) is modeled as a chain of beads of diameter b connected by springs having center-to-center equilibrium distance $b(1 + \varepsilon)$, with $\varepsilon > 0$. The numerical value of b is obtained

---

[2]By definition, 1 M=1 mol/l [74].
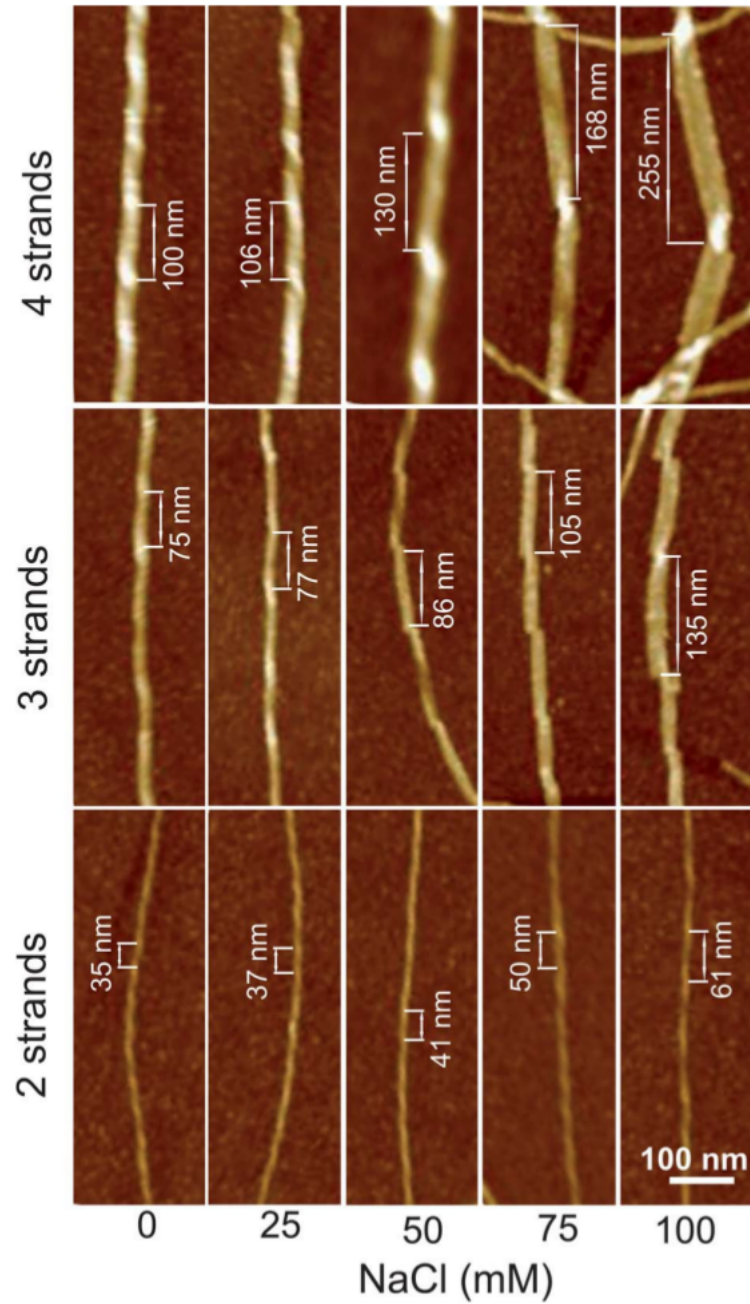
Figure 1.8 – Sample AFM images and corresponding halfpitches for fibrils with $n = 2, 3, 4$ fibrils at $0, 25, 50, 75, 100$ mM salt concentrations. Reprinted from [44].

directly from experiments as the height of individually adsorbed protofilaments, and is set to 2 nm [43]. Within our coarse-grained picture, the springs embody the elastic part of the energy, while the electrostatic interaction is accounted for by assigning a charge $z \cdot e$ to each bead (e being the electron charge, and z the number of charges on the bead). A n-stranded ribbon is
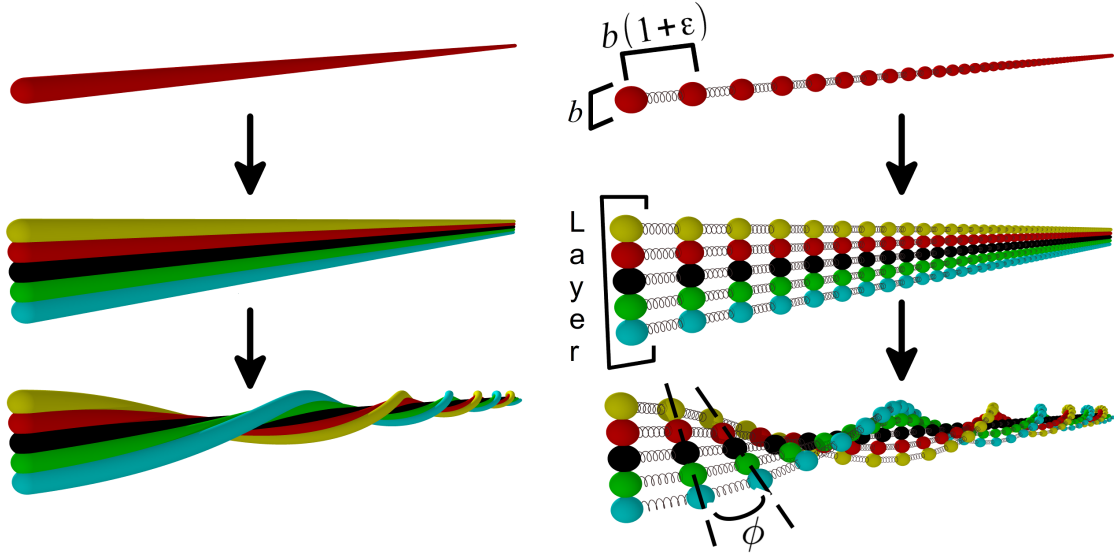
Figure 1.9 – Ideal steps leading to the coarse-grained picture of a fibril. A single protofilament (top-left) is represented by a chain of beads (top-right). Several pre-formed protofilaments associate laterally (medium-left) and the whole structure is subsequently twisted around its longitudinal axis (bottom-left), thus obtaining the final fibril. Accordingly, in our model a suitable number of chains are attached to each other (medium-right) and homogeneously twisted (bottom-right), with consecutive transverse layers tilted by a certain angle ϕ. Reprinted from [67].

formed by laterally associating n chains, so that transverse layers of beads are formed (one such layer is highlighted in fig.1.9). The final structure is then obtained by twisting the ribbon around its longitudinal axis, as shown in fig.1.9 for a $n = 5$ fiber (the final structures for $n = 2, 3$ and 4 are also shown in fig.1.10). In what follows we assume each layer not to bend upon twisting, so to store elastic energy primarily by torsion and not bending. This assumption of perfectly-straight fibril axis is justified because the persistence lengths of fibrils are much larger than the halfpitches (for example, comparing the data shown in fig.1.6 we see that l_p is at least 30 times larger than L). Moreover, in the present model only interactions between consecutive beads within the same protofilament are taken into account. This is a reasonable assumption because b is of the same order of the typical values of the screening length in aqueous solutions [71], so that electrostatic interactions are negligible between non-adjacent layers.

Because of the geometrical constraints and of the simultaneous action of the two competing forces, we expect at equilibrium that two consecutive layers will be tilted with respect to each other by a certain angle ϕ^*, which is assumed to be constant along the ribbon and small. Since all energy terms in our derivation do not prefer any particular local handedness of the twist, the interaction energy is an even function of the twist angle, so that its expected shape is that of a double well, which is sketched in fig.1.11a. Nevertheless, the fibrils do show a uniform
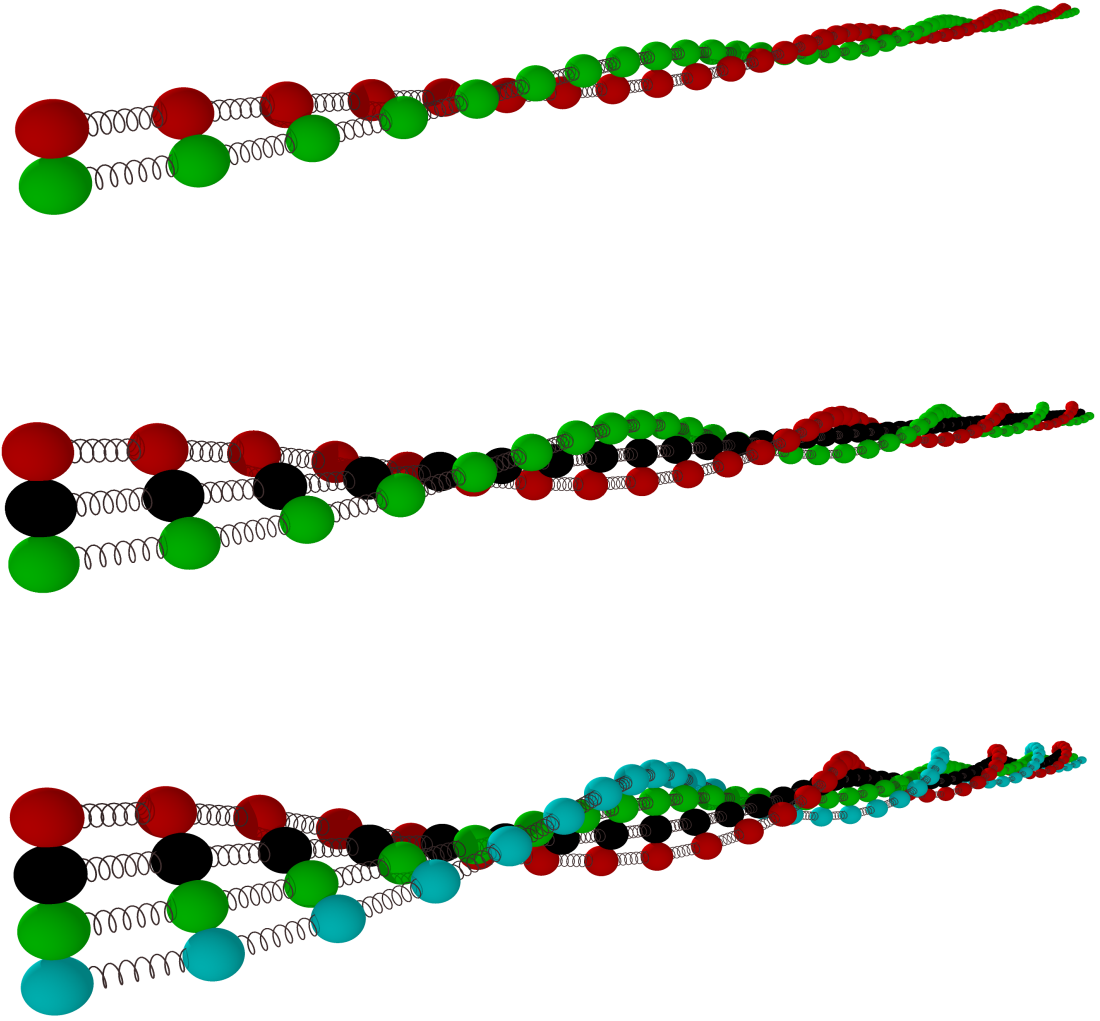
Figure 1.10 – Final structures obtained by following the procedure described in the text for fibrils made of 2, 3 and 4 filaments. Reprinted from [67].

left handedness over the full fibril length [43, 44], which means that a symmetry-breaking mechanism is at work, imposing an effective chirality on the system. In what follows, we postulate this intrinsic feature to be responsible for giving a definite handedness, but to play a negligible role for all other purposes.

1.3.2 Computation of the equilibrium angle

The (still unknown) small value of ϕ^* allows to set up a perturbative scheme to compute the energy of the full fibril, which eventually leads to a very powerful formula for the half-pitch L. The present section is devoted to the computation of ϕ^*, thus it may be skipped if the reader is not interested in the detailed calculations.
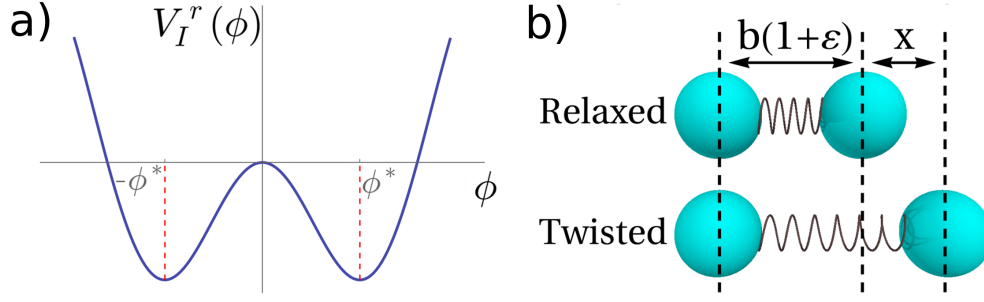
Figure 1.11 – a) due to the competition between elasticity and the repulsive electrostatic interaction, the interaction energy between two consecutive layers as a function of ϕ has a double-well shape, where the minima $\pm\phi^*$ are obtained as a compromise between a perfectly relaxed ($\phi = 0$) and a strongly twisted ($\phi = \pi/2$) conformation. b) the variable x is defined as the displacement of the distance between two consecutive beads with respect to the untwisted case. Adapted from [67].

Let us consider two consecutive layers on the fibril. As mentioned above, assuming chirality not to play a relevant role in quantitatively determining the twist of the ribbon, the total (elastic + electrostatic) interaction energy between two consecutive layers $V_I^r(\phi)$ (where the subscript stresses its dependence on the ionic strength I of the solution) is expected to be an even function of the angle ϕ, so that the point $\phi = 0$ is stationary (*i.e.* maximum or minimum). However, such an equilibrium point corresponds to an unstable configuration. Indeed, the $\phi = 0$ state coincides with a perfectly relaxed ribbon, which must still choose the twisting handedness as driven by chirality. Now, let $U_I(x)$ be the total interaction energy between two beads having a distance equal to $x + b(1 + \varepsilon)$ (we will refer to x as *displacement*, see fig.1.11b). If we denote by $x_i(\phi)$ the displacement of the distance between two beads on the ith protofilament belonging to two consecutive layers tilted by ϕ, by definition we have

$$V_I^r(\phi) = \sum_i U_I(x_i(\phi)). \tag{1.9}$$

Since we assumed ϕ to be small, the distance x_i is expected to be very small when compared to the spring equilibrium distance $b(1 + \varepsilon)$: $x_i(\phi) \ll b(1 + \varepsilon)$. Therefore, we can perform a second-order Taylor expansion of the right-hand side of equation (1.9), thus obtaining (up to an additive term independent of ϕ)

$$V_I^r(\phi) \simeq U_I'(0) \sum_i x_i(\phi) + \frac{1}{2} U_I''(0) \sum_i x_i^2(\phi). \tag{1.10}$$

In order to compute the displacement $x_i(\phi)$, we first need to address the calculation of the center-to-center distance $d_i(\phi)$ between the centers of two interacting beads on the ith filament, which by definition satisfies $d_i(\phi) \equiv x_i(\phi) + b(1 + \varepsilon)$. In order to do so, let us consider two consecutive layers on the fibril. By construction, such layers are tilted with respect to

each other by an angle ϕ. Let us consider the reference frame having as x axis the axis of the fibril and as y axis the direction of the first layer (*i.e.* the line connecting the centers of the beads belonging to it). Finally, let the z axis be suitably chosen in order to obtain a left-handed orthonormal reference frame. If we consider a bead belonging to the first layer, by construction its x and z coordinates are equal to zero. Moreover, apart from a sign the y coordinate corresponds to its distance with respect to the fibril axis. If we label the filaments as $-(n-1)/2, -(n-1)/2+1, \ldots, (n-1)/2$ starting from the filament more shifted towards the negative values of y, we then easily obtain that the y coordinate of the bead belonging to the ith filament is simply ib, where b is its diameter. As for the second layer, by construction all its beads have $x = b(1+\varepsilon)$. As for the y and z coordinates, by following the same reasoning as above and remembering that the second layer is rotated around the x axis by ϕ, we find $y = ib\cos\phi, z = ib\sin\phi$. Putting together the previous results, we can thus straightforwardly compute the distance $d_i(\phi)$ separating two consecutive beads belonging to the ith filament as

$$d_i(\phi) = \sqrt{(b(1+\varepsilon) - 0)^2 + (ib - ib\cos\phi)^2 + (0 - ib\sin\phi)^2}, \tag{1.11}$$

which gives for the displacement

$$x_i(\phi) = b\sqrt{(1+\varepsilon)^2 + 2i^2 - 2i^2\cos\phi} - b(1+\varepsilon). \tag{1.12}$$

By Taylor expanding $x_i(\phi)$, we finally find

$$x_i(\phi) = \alpha i^2 \phi^2 + \beta i^2 \phi^4 + \gamma i^4 \phi^4 + O(\phi^6), \tag{1.13}$$

where $\alpha = b/[2(1+\varepsilon)], \beta = -b/[24(1+\varepsilon)]$ and $\gamma = -b/[8(1+\varepsilon)^3]$.

Substituting the previous result into equation (1.10) and neglecting all the terms $O(\phi^6)$, we find

$$V_I^r(\phi) \simeq U_I'(0) \sum_i \left(\alpha i^2 \phi^2 + \beta i^2 \phi^4 + \gamma i^4 \phi^4\right) + \frac{1}{2} U_I''(0) \sum_i \alpha^2 i^4 \phi^4, \tag{1.14}$$

that is, by rearranging

$$V_I^r(\phi) \simeq U_I'(0) \left(\alpha\phi^2 + \beta\phi^4\right) \sum_i i^2 + \left(U_I'(0)\gamma + \frac{1}{2} U_I''(0)\alpha^2\right) \phi^4 \sum_i i^4. \tag{1.15}$$

By making use of the formulas

$$\sum_{i=-\frac{n-1}{2}}^{\frac{n-1}{2}} i^2 = \frac{1}{12} n(n^2 - 1) \tag{1.16}$$

$$\sum_{i=-\frac{n-1}{2}}^{\frac{n-1}{2}} i^4 = \frac{1}{240} n(n^2 - 1)(3n^2 - 7) \tag{1.17}$$

the previous expression becomes

$$V_I^r(\phi) \simeq U_I'(0)\left(\alpha\phi^2 + \beta\phi^4\right)\frac{n(n^2-1)}{12} + \left(U_I'(0)\gamma + \frac{1}{2}U_I''(0)\alpha^2\right)\phi^4 \frac{n(n^2-1)(3n^2-7)}{240} =$$

$$= \frac{1}{12}n(n^2-1)\left\{U_I'(0)\alpha\phi^2 + \left[U_I'(0)\beta + \frac{3n^2-7}{20}\left(U_I'(0)\gamma + \frac{1}{2}U_I''(0)\alpha^2\right)\right]\phi^4\right\}. \tag{1.18}$$

As we anticipated, within our assumptions the potential is an even function of the angle ϕ. By minimizing the previous expression with respect to ϕ, we end up with the following equation:

$$2U_I'(0)\alpha\phi + 4\left[U_I'(0)\beta + \frac{3n^2-7}{20}\left(U_I'(0)\gamma + \frac{1}{2}U_I''(0)\alpha^2\right)\right]\phi^3 = 0. \tag{1.19}$$

A trivial solution is obviously $\phi = 0$. However, we note that, because of the simultaneous presence of the attractive elastic term and the electrostatic repulsion, the function $U_I(x)$ is expected to present a single minimum for some $0 < x \ll b(1+\varepsilon)$, so that $U_I'(0) < 0$ and $U_I''(0) > 0$. Since $U_I'(0) < 0$ we easily obtain

$$\frac{d^2 V_I^r}{d\phi^2}(0) = \frac{1}{6}n(n^2-1)U_I'(0)\alpha < 0, \tag{1.20}$$

so that this point corresponds to a maximum, as we had already anticipated by physical arguments. As for the other two solutions, because of the parity of the potential they have the same absolute value ϕ^* given by

$$\frac{1}{\phi^*} = \sqrt{-\frac{2\left[U_I'(0)\beta + \frac{1}{20}(3n^2-7)\left(U_I'(0)\gamma + \frac{1}{2}U_I''(0)\alpha^2\right)\right]}{U_I'(0)\alpha}}. \tag{1.21}$$

By substituting $\alpha = b/[2(1+\varepsilon)], \beta = -b/[24(1+\varepsilon)], \gamma = -b/[8(1+\varepsilon)^3]$ and rearranging, we thus find

$$\frac{1}{\phi^*} = \sqrt{\frac{1}{6} + \frac{3n^2-7}{40(1+\varepsilon)^2} - \frac{3n^2-7}{40}\frac{U_I''(0)}{U_I'(0)}\frac{b}{1+\varepsilon}}. \tag{1.22}$$

An explicit computation of the second derivative in these points confirms that they are minima of the potential $V_I^r(\phi)$. The stable equilibrium point is given by the minimum corresponding to the correct chirality.

Since we assumed $\phi \ll \pi/2$, we expect the right-hand side of the previous formula to be much greater than one. However, the first two terms in the square root do not justify $1/\phi^* \gg 1$ (remember that n is on the order of unity and $\varepsilon > 0$), so that we expect the third one to be the leading term. The final formula for the equilibrium angle is thus (below we will countercheck the correctness of the previous statement by computing the neglected terms within the explicit

model)

$$\frac{1}{\phi^*} = \sqrt{-\frac{3n^2 - 7}{40} \frac{U_I''(0)}{U_I'(0)} \frac{b}{1+\varepsilon}}.$$

(1.23)

1.3.3 A universal blueprint in the twisting properties of fibrils

The halfpitch L of the fibril can be obtained by computing how many layers are necessary to accumulate a π twisting angle. Since the distance between two consecutive layers is set to $b(1+\varepsilon)$, the proportion $L:\pi = b(1+\varepsilon):\phi^*$ holds. By substituting equation (1.23), we thus find

$$L(n,I) = \pi b \sqrt{-\frac{3n^2 - 7}{40} \frac{U_I''(0)}{U_I'(0)} b(1+\varepsilon)}.$$

(1.24)

The previous formula has some remarkable features. Indeed, we note that in our model the population to which the particular fibril under study belongs to is identified by means of the variable n. Intriguingly, equation (1.24) depends on n by means of a multiplicative factor which is completelely independent of the other relevant quantities of the system. In other words, our model leads to a net separation of the mesoscale properties of the system (that is, the n-dependence) and its microscopic details, such as the ionic strength I of the solution, the elastic properties, and the basic geometrical features (b and ε) at the level of single beads. Apart from being physically sound, this observation leads to some new important insights in the mesoscopic structure of fibrils. Indeed, because the details of the electrostatic and elastic interactions are concealed in the multiplicative n-independent terms in equation (1.24), the ratio between halfpitches of fibrils composed of different numbers n_2 and n_1 of protofilaments depends only on n_1 and n_2:

$$\frac{L(n_2, I)}{L(n_1, I)} = \sqrt{\frac{3n_2^2 - 7}{3n_1^2 - 7}},$$

(1.25)

which is a remarkable parameter-free prediction. We leave to later the discussion on the explicit form of the interaction potential, and turn first to the noteworthy predictive power of the $L(n_2, I)/L(n_1, I)$ expression in equation (1.25).

In order to benchmark our predictions against experimental data, in collaboration with the group of Prof. Mezzenga at ETHZ we considered a set of βlg fibrils produced with the same protocol as in the previous sections. After the aggregation process, fibrils were immersed in solutions with different salinities as in [44] and let equilibrate for one year. When deposited on mica, they displayed periodicities in quantitative agreement with the experimental values in [44]. Since the latter were performed on fibrils which had rested in salty solutions for three weeks, this agreement confirms that the structures we observe correspond to the thermodynamic equilibrium. Moreover, the statistical populations of wide fibrils were rich enough to

enable us obtain data on fibrils made of up to 5 protofilaments for all the ionic strength values considered ($I = 10, 35, 60, 85$ and 110 mM). Furthermore, in order to rule out strong effects due to the interaction with the substrate, the same measurements were repeated on fibrils adsorbed on HOPG. While the interaction responsible for adsorption on mica is mainly of electrostatic nature, HOPG is a strongly hydrophobic surface. Therefore, the observation of compatible values for the halfpitches between the two cases suggests that substrates effects do not strongly alter the bulk periodic features of the fibrils (nevertheless, as we show below, by addressing these effects even at the simplest level leads to an even better quantitative agreement between theory and experiments).
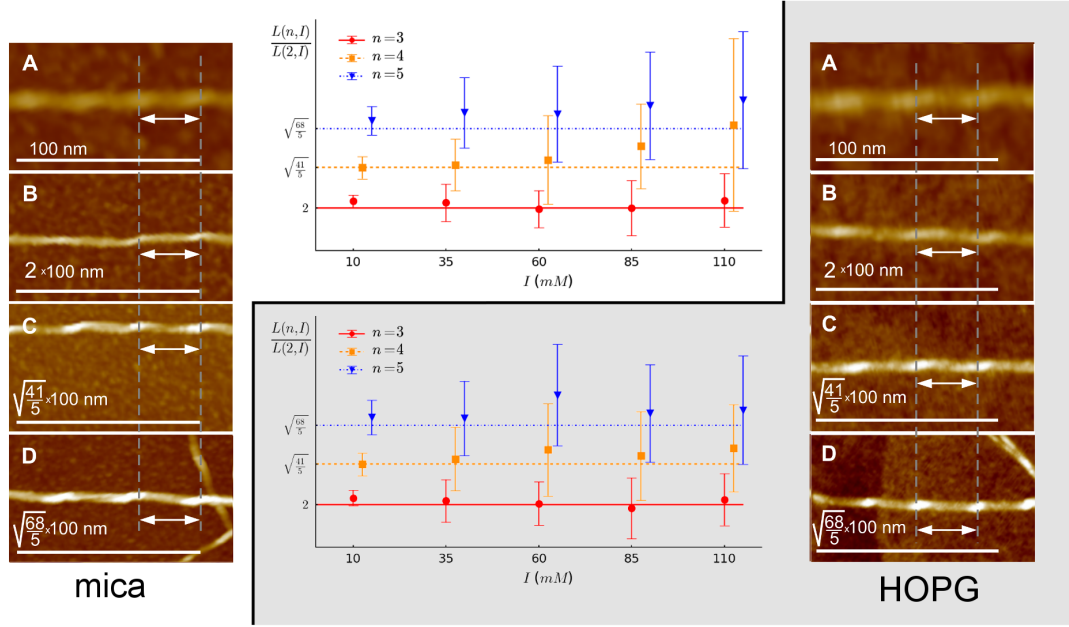
Figure 1.12 – AFM images of fibrils composed of a different number of protofilaments (A,B,C,D correspond respectively to $n = 2, 3, 4, 5$), incubated on a 10 mM solution and adsorbed on mica (left image) and HOPG (right image) surfaces. By performing a proper rescaling of the scale-bar according to the formula given in the main text, for both substrates the half-pitches appear to have the same physical length. In the central column, we compare the experimental values of the ratios $L(n, I)/L(2, I)$ for $n = 3, 4, 5$ (points) with the corresponding theoretical predictions (straight lines) for several values of ionic strength I in the case of both mica and HOPG substrates. Reprinted from [67].

In fig.1.12 we show AFM mosaic images giving the half-pitch of multistranded βlg fibrils with $n = 2, 3, 4$ and 5 constitutive protofilaments without added salt (ionic strength of pH 2 pure water, *i.e.* 10 mM) acquired on mica (left image) and HOPG (right image) substrates. The scale bar has identical physical length in each case, but indicates rescaled dimensions of 100 nm x $(3n_2^2 - 7)^{1/2}/(3n_1^2 - 7)^{1/2}$ for each multistranded case, respectively n_2 being the number of protofilaments composing the fibril under inspection and $n_1 = 2$ serving as a reference. After

operating the above-mentioned rescaling, all the pitches become identical, consistently with our predictions. Because the predicted behavior is universal, *i.e.* independent of the physical parameters and in particular of the ionic strength, in fig.1.12 we further show the comparison between theory and experiments at different ionic strengths (central column). Particularly, we plot $L(n, I)/L(2, I)$ for the sets of data for the two surfaces (data corresponding to the various values of n have been horizontally shifted in order to improve visualization), and horizontal lines correspond to our theory. The agreement between theory and experimental data is very good, given the total absence of fitting parameters in the predictions.
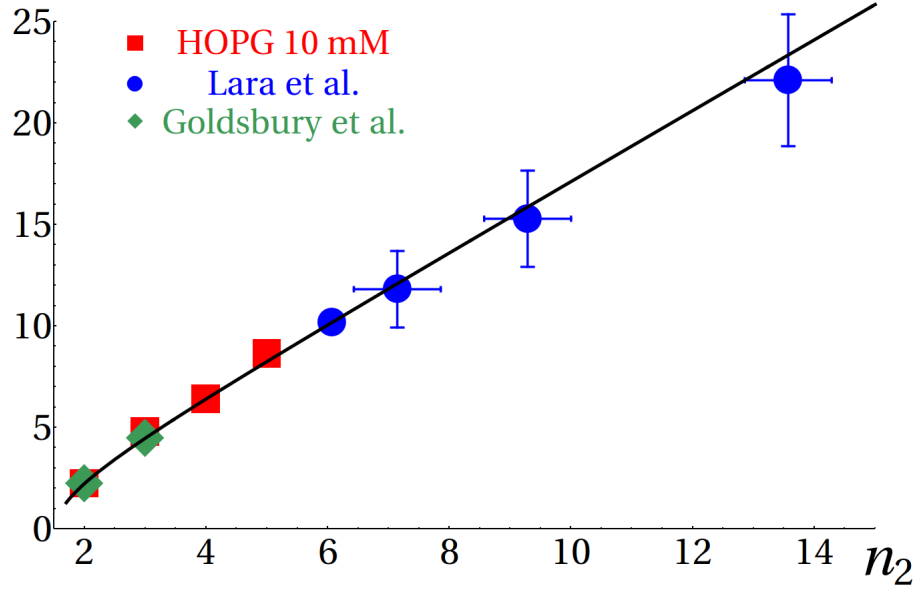
Figure 1.13 – Universal prediction vs experimental data. In this graph we plot the rescaled ratio $\sqrt{3n_1^2 - 7}L(n_2)/L(n_1)$ of the periodicities of fibrils made from three different proteins, but all arranged in a twisting ribbon: i) a set of βlg fibrils considered in the present work ($I = 10$ mM), whose periodicity was measured by means of AFM and using HOPG as substrate (red squares, $n_1 = 2$); ii) cryo-TEM data on lysozyme fibrils from [75] (blue circles, $n_1 \simeq 6$); and iii) TEM data of human amylin fibrils from [76] (green diamonds, $n_1 = 2$). The data nicely collapse on the curve $\sqrt{3n_2^2 - 7}$ (solid line) as predicted by our model. Reprinted from [67].

The level of coarse grain of our model and the separation of scales in the formula for the halfpitch make the prediction given by equation (1.25) very general. Indeed, no particular information of the microscopic details of βlg fibrils is needed in order to arrive at this formula. Therefore, one may wonder whether it is applicable also to other systems, provided that their geometry is that of a flat ribbon. Strikingly, this is exactly the case. Particularly, we now focus on two other ribbon-shaped amyloid systems studied in literature, namely lysozime fibrils, which were studied by means of AFM and cryo-TEM in [75], and human amylin fibrils, characterized by TEM in [76]. By rearranging equation (1.25), independently of the microscopic details of

the system and the reference value n_1, the following relation holds:

$$\frac{L(n_2)}{L(n_1)}\sqrt{3n_1^2 - 7} = \sqrt{3n_2^2 - 7}. \tag{1.26}$$

In other words, by multiplying the ratio of the half-pitches by $\sqrt{3n_1^2 - 7}$, where n_1 is the number of protofilaments composing the reference fibril, according to our model all the data collapse on the universal curve $\sqrt{3n_2^2 - 7}$. In fig.1.13, we report the results found by considering the data in [75] (blue circles) and [76] (green diamonds), together with a set of data considered in the present work (red squares). The agreement with our parameter-free law (solid line) is strikingly good, and confirms the goodness of our approach in describing different systems arranged in the same geometrical packing.

1.3.3.1   Different geometries result into new universal laws
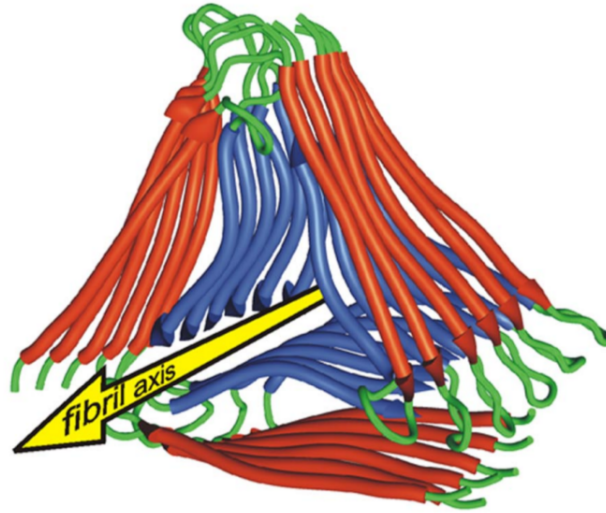
Figure 1.14 – Threefold structure proposed as possible packing arrangement for Aβ_{1-40} amyloid fibrils [59]. Each protofilament is obtained by stacking onto each other the Aβ_{1-40} peptides, represented in the picture as the C-shaped units formed by a red and a blue β strands connected by a green linker. The three protofilaments are then arranged in a triangular fashion and twist around each other. Reprinted from [59].

Equation (1.25) was derived by considering a flat-ribbon packing arrangement. Naturally, this is not the only possible geometry for fibrils. For example, based on STEM and NMR data, it was proposed [59] that Aβ_{1-40} amyloid fibrils can exist in polymorphic states, which include ribbon-like arrangements and the threefold symmetric structure shown in fig.1.14. Though our calculations have been performed for the specific geometry of a flat ribbon, the approach can be straightforwardly generalized to other packing arrangements. By following the same steps performed to find equation (1.25), we now derive as an example a new parameter-free

law in the case of protofilaments arranged in a twisting hollow-tube geometry, *i.e.* when in the section of a fibril made of n protofilaments the latter are located on a regular n-sided polygon. For example, in the top panels of fig.1.15 we show the cartoons of two fibrils packed according to this geometry with $n = 3$ (which corresponds to the model for $A\beta_{1-40}$ mentioned above) and $n = 5$.
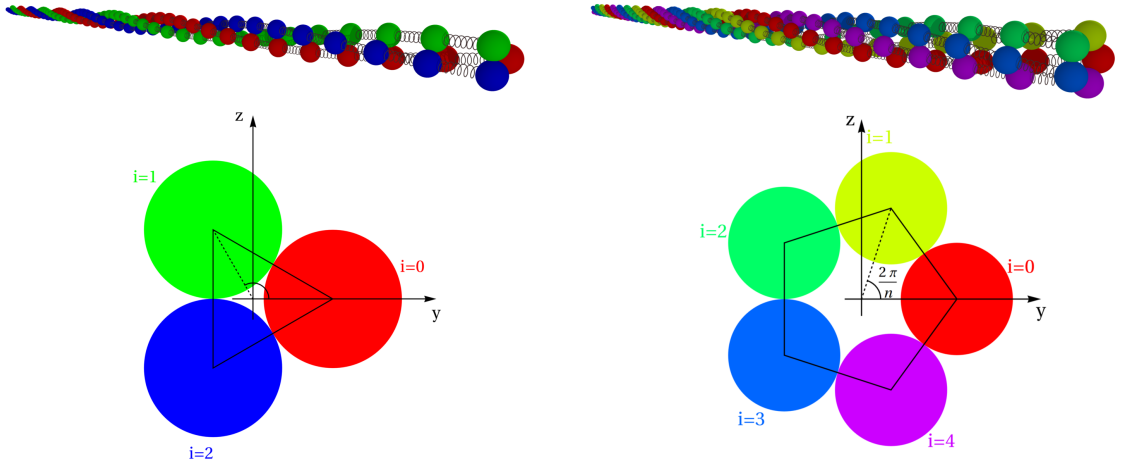
Figure 1.15 – Examples of twisting hollow-tube geometry. Left: cartoon of a twisting fibril made of $n = 3$ protofilaments arranged on an equilateral triangle (top) and basic geometrical features of a layer (bottom). Right: as in left column, but with $n = 5$. In both cases, the arbitrary protofilament chosen to fix the y axis (corresponding to $i = 0$) is depicted in red. Reprinted from [67].

By proceeding as in section 1.3.2, we focus on two consecutive layers on the fibril. By construction, such layers are tilted with respect to each other by an angle ϕ. Let us consider the reference frame having as x axis the axis of the fibril and as y axis the line connecting the center of the polygon to the center of an arbitrarily selected bead (bottom panels of fig.1.15). Finally, let the z axis be suitably chosen in order to obtain a left-handed orthonormal reference frame. Within this arrangement, let us label the protofilaments as $0, \ldots, n-1$ starting from the one chosen to define the y axis. If we consider a bead belonging to the first layer, by construction its x coordinate is equal to $x_1 = 0$, while the other two coordinates can be found by simple trigonometric considerations to be equal to

$$y_1 = \frac{b}{2\sin\frac{\pi}{n}}\cos\frac{2\pi i}{n} \qquad z_1 = \frac{b}{2\sin\frac{\pi}{n}}\sin\frac{2\pi i}{n}. \qquad (1.27)$$

Indeed, the side of the polygon corresponds to the center-to-center distance b between beads belonging to two consecutive protofilaments, while the distance separating any bead from the center of the polygon corresponds to the radius of the circumscribed circle of the polygon, and is equal to $b/2\sin(\pi/n)$.

29

As for the second layer, its x coordinate is equal to $x_2 = b(1 + \varepsilon)$, while the y and z coordinates can be found rotating by ϕ the ones found for the first layer:

$$y_2 = \frac{b}{2 \sin \frac{\pi}{n}} \cos \left(\frac{2 \pi i}{n} + \phi \right) \qquad z_2 = \frac{b}{2 \sin \frac{\pi}{n}} \sin \left(\frac{2 \pi i}{n} + \phi \right). \tag{1.28}$$

The distance d_i between two consecutive beads belonging to the i-th protofilament is then equal to

$$d_i = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}, \tag{1.29}$$

which results, after some manipulations, in

$$d_i(\phi) = b \sqrt{(1 + \varepsilon)^2 + \frac{1 - \cos \phi}{2 \sin^2 \frac{\pi}{n}}}. \tag{1.30}$$

As expected from the symmetry of the system, the distance d_i is independent of i. At this point, we can Taylor-expand $x_i = d_i - b(1 + \varepsilon)$ and find

$$x_i(\phi) = x(\phi) = \frac{b}{8(1 + \varepsilon)} \frac{\phi^2}{\sin^2 \frac{\pi}{n}} - \frac{b}{96(1 + \varepsilon)} \frac{\phi^4}{\sin^2 \frac{\pi}{n}} - \frac{b}{128(1 + \varepsilon)^3} \frac{\phi^4}{\sin^4 \frac{\pi}{n}} + O(\phi^6), \tag{1.31}$$

where by introducing $x(\phi)$ we wanted to stress the independence of x_i on i. The total energy of a layer V_I^h can be computed by summing the contributions coming from each filament:

$$V_I^h(\phi) \simeq U_I'(0) \sum_{i=0}^{n-1} x_i(\phi) + \frac{1}{2} U_I''(0) \sum_{i=0}^{n-1} x_i^2(\phi) = n U_I'(0) x(\phi) + \frac{1}{2} n U_I''(0) x^2(\phi), \tag{1.32}$$

that is, after substituting the formula for $x(\phi)$, retaining only the terms up to ϕ^4 and rearranging,

$$V_I^h(\phi) \simeq \frac{b}{8(1 + \varepsilon)} \frac{n}{\sin^2 \frac{\pi}{n}} \left\{ U_I'(0) \phi^2 + \left[\frac{U_I''(0) b}{16(1 + \varepsilon) \sin^2 \frac{\pi}{n}} - \frac{U_I'(0)}{12} - \frac{U_I'(0)}{16(1 + \varepsilon)^2 \sin^2 \frac{\pi}{n}} \right] \phi^4 \right\}. \tag{1.33}$$

By minimizing V_I^h with respect to ϕ, we find the equilibrium twisting angle ϕ^*:

$$\frac{1}{\phi^*} = \sqrt{\frac{1}{6} + \frac{1}{8(1 + \varepsilon)^2 \sin^2 \frac{\pi}{n}} - \frac{b}{8(1 + \varepsilon)} \frac{U_I''(0)}{U_I'(0)} \frac{1}{\sin^2 \frac{\pi}{n}}}. \tag{1.34}$$

Arguments similar to the ones used in section 1.3.2 lead to recognize the third term in the square root of the previous expression as the leading one, thus giving as a final formula for ϕ^*

$$\frac{1}{\phi^*} = \frac{\sqrt{-\frac{b}{8(1 + \varepsilon)} \frac{U_I''(0)}{U_I'(0)}}}{\sin \frac{\pi}{n}} \tag{1.35}$$

which results into a half-pitch equal to

$$L(n, I) = \pi b (1 + \varepsilon) \frac{\sqrt{-\frac{b}{8(1+\varepsilon)} \frac{U_I''(0)}{U_I'(0)}}}{\sin \frac{\pi}{n}} \ . \tag{1.36}$$

Exactly as in the case of a ribbon, we can recognize two multiplicative terms coming from the micro- and mesoscopic scales. This results into a new parameter-free law, qualitatively different from equation (1.25), which holds for a hollow-tube geometry: taking fibrils of different populations (*i.e.* different n) and computing the ratio of the half-pitches, we find

$$\frac{L(n_2, I)}{L(n_1, I)} = \frac{\sin \frac{\pi}{n_1}}{\sin \frac{\pi}{n_2}} \ . \tag{1.37}$$

Unfortunately, the packing arrangement presented in [59] cannot be used to directly test this prediction, since at least two distinct architectures are needed in order to compute this ratio. The present derivation has thus to be intended only as an example to show how different arrangements result into new parameter-free laws.

1.4 Explicit model for βlg fibrils

1.4.1 An explicit formula for the halfpitch

Having established the correctness and universality of our findings, we turn to the explicit form of the interaction energy $U_I(x)$. The elastic contribution to the energy is captured by Hookean springs between consecutive beads, with spring constant k_e. To estimate the electrostatic part of the energy, we use the Poisson-Boltzmann approximation, where the electrostatic interaction energy between two spherical surfaces of diameter b and charge $z \cdot e$ at a minimum approach distance x, immersed in a solution of ionic strength I, can be written as [71]

$$U_I^{\text{electrostatic}} = \frac{1}{4} b Z(I) e^{-k(I)x} \ . \tag{1.38}$$

Here, $k(I)$ is the inverse of the Debye length, which describes how well the ions in solution screen the Coulomb repulsion, and is equal to

$$k(I) = \sqrt{\frac{2 N_A e^2 I}{\epsilon_0 \epsilon_r k_B T}} \ , \tag{1.39}$$

where $\epsilon_r \simeq 78$ is the dielectric constant of water and I the ionic strength of the solution measured in mM. Moreover, apart from the Yukawa-like exponential term, the electrostatic potential in equation (1.38) is further screened by the I-dependent interaction parameter Z,

which is equal to

$$Z(I) = 64\pi\epsilon_0\epsilon_r \left(\frac{k_B T}{e}\right)^2 \tanh^2 \frac{e\psi_0(I)}{4k_B T}. \tag{1.40}$$

In this formula, $\psi_0(I)$ is the electrostatic potential generated by a charged sphere on its surface, and has to be determined by solving the equation $\sigma_0 = \sqrt{8\epsilon_0\epsilon_r k_B T N_A I} \sinh\frac{e\psi_0(I)}{2k_B T}$, where $\sigma_0 = ze/\pi b^2$ is the surface charge density. The previous formulas can be written in a simpler form by considering the *Bjerrum length*, which by definition is equal to the distance at which two elementary charges have an interaction energy equal to $k_B T$ and has formula

$$l_B \equiv \frac{e^2}{4\pi\epsilon_0\epsilon_r k_B T}. \tag{1.41}$$

In water at room temperature the Bjerrum length has a value roughly equal to 7 Å. Plugging in the expression of l_B into the previous formulas we easily find

$$k(I) = \sqrt{8\pi N_A l_B I} \tag{1.42}$$

and

$$Z(I) = \frac{16 k_B T}{l_B} \tanh^2 \frac{e\psi_0(I)}{4k_B T}, \tag{1.43}$$

while the equation defining $\psi_0(I)$ now reads $\sqrt{2\pi N_A I/l_B} \sinh(e\psi_0(I)/2k_B T) = z/b^2$, where we have substituted the formula for σ_0. Note that the interior of the fibrils is essentially dry, and that it is mostly through the exterior that fibrils interact via a screened electrostatic potential. Thus, equation (1.38) approximates the screened electrostatic interaction between nearby portions of the fibril surface, in order to provide a coarse-grained picture of the twisting properties of fibrils.

The total interaction energy between two consecutive beads on a given protofilament can be thus written as

$$U_I(x) = \frac{1}{2} k_e x^2 + \frac{1}{4} b Z(I) e^{-k(I)(x+b\varepsilon)}. \tag{1.44}$$

By computing the derivatives of $U_I(x)$ we easily find (to ease the notation we drop out the I dependence in $k(I)$ and $Z(I)$)

$$U_I'(x) = k_e x - \frac{1}{4} k b Z e^{-k(x+b\varepsilon)}, \tag{1.45}$$

$$U_I''(x) = k_e + \frac{1}{4} k^2 b Z e^{-k(x+b\varepsilon)}. \tag{1.46}$$

In this way, we have

$$U_I'(0) = -\frac{1}{4}kbZe^{-kb\varepsilon}, \tag{1.47}$$

$$U_I''(0) = k_e + \frac{1}{4}k^2 bZe^{-kb\varepsilon}. \tag{1.48}$$

As expected, $U_I'(0) < 0$ and $U_I''(0) > 0$. Substituting in equation (1.23), we obtain

$$\frac{1}{\phi^*} = \sqrt{\frac{3n^2-7}{40}\frac{k_e + \frac{1}{4}k^2 bZe^{-kb\varepsilon}}{\frac{1}{4}kbZe^{-kb\varepsilon}}\frac{b}{1+\varepsilon}} = \sqrt{\frac{3n^2-7}{10}\frac{k_e e^{kb\varepsilon}}{kZ(1+\varepsilon)} + \frac{3n^2-7}{40}\frac{kb}{1+\varepsilon}}. \tag{1.49}$$

Now, in the range of I values considered in the experiment, the largest value of kb is attained for the largest concentration of counterions (equal to 110 mM) and is roughly equal to 2.2. By following the same reasoning as in section 1.3.2, we can thus neglect the second term within the square root, obtaining as a final formula

$$\frac{1}{\phi^*} = \sqrt{\frac{3n^2-7}{10}\frac{k_e e^{kb\varepsilon}}{kZ(1+\varepsilon)}}. \tag{1.50}$$

Thus, plugging this expression in equation (1.24), we finally find

$$L(n,I) = \pi b\sqrt{\frac{3n^2-7}{10}\frac{k_e e^{kb\varepsilon}}{kZ}(1+\varepsilon)}. \tag{1.51}$$

1.4.2 Fitting the raw data

Many of the parameters present (implicitly or explicitly) in equation (1.51) can be directly determined from the experiments. The ones making exception are the empirical parameters k_e and ε, which describe the elastic part of the potential, and the total charge z present on each bead. By tuning the parameters k_e and ε, in the present section we fit the raw experimental values of the halfpitches by means of equation (1.51). Note that for each substrate all the 20 experimental data points are used simultaneously in the fit (i.e. for each substrate a single pair of $k_e : \varepsilon$ allows fitting simultaneously all the 20 experimental observations), thus as we show below a single choice of k_e and ε is capable of quantitatively describing the twisting periodicities of the fibrils.

Before performing the fit, we need to deal with the total charge of a bead z. Since in the native state a βlg unit has a roughly spherical shape of diameter 2 nm, one could assume that each bead corresponds to roughly one monomer, and consequently assign to it a charge $z = 20$ [43]. However, physical arguments based on Manning theory of counterion condensation [77] point towards $z = 6$ as the most suitable value. Indeed, after the aggregation has taken place, a value
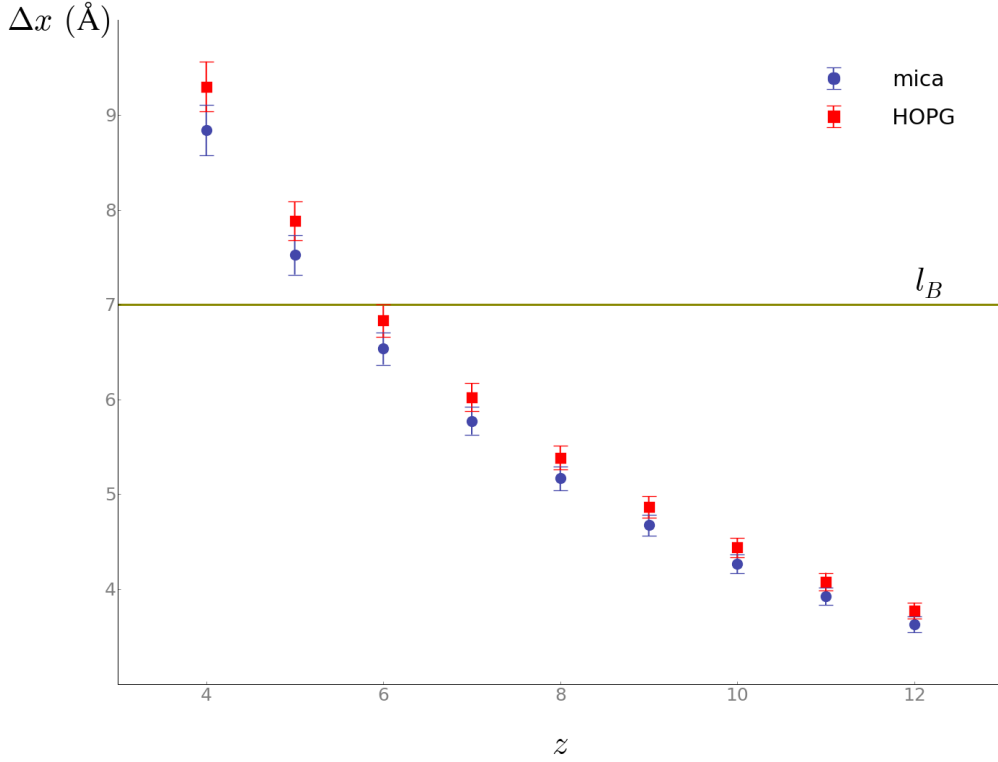
Figure 1.16 – Mean spacing between two unit charges on each filament as a function of the bead charge. For any given value of z, we performed the fit and computed $\Delta x = b(1+\varepsilon)/z$ by substituting the given value of ε found. The blue circles and red squares correspond respectively to the data obtained on mica and HOPG. The horizontal line sketched corresponds to the Bjerrum length, so that the chosen z is the one for which the corresponding Δx lies closest to it. Reprinted from [67].

of $z = 20$ would result in a densely charged filament, which according to Manning's theory [77] would result in a partial condensation of counterions on it. More specifically, within our model the average spacing between two unit charges is given by $\Delta x \equiv b(1+\varepsilon)/z$. Given that $b = 2$ nm and the fact that ε has always a value around unity (see below), a value of $z = 20$ would imply a mean spacing $\Delta x \simeq 2$ Å. According to Manning's theory, the counterion condensation takes place whenever the mean spacing falls below the Bjerrum length l_B, which in the experimental conditions considered in this work is equal to roughly 7 Å. The theory further states that such condensation is only partial, and involves a suitable number of counterions in order to obtain a mean spacing equal to l_B after it has taken place.

Based on the previous remarks, we can thus conclude that the more physically-reliable z is the one which results in a mean spacing $\Delta x \simeq l_B$. However, since Δx depends on ε, its value can be determined only *a posteriori* with respect to the fit. We thus performed the fit of
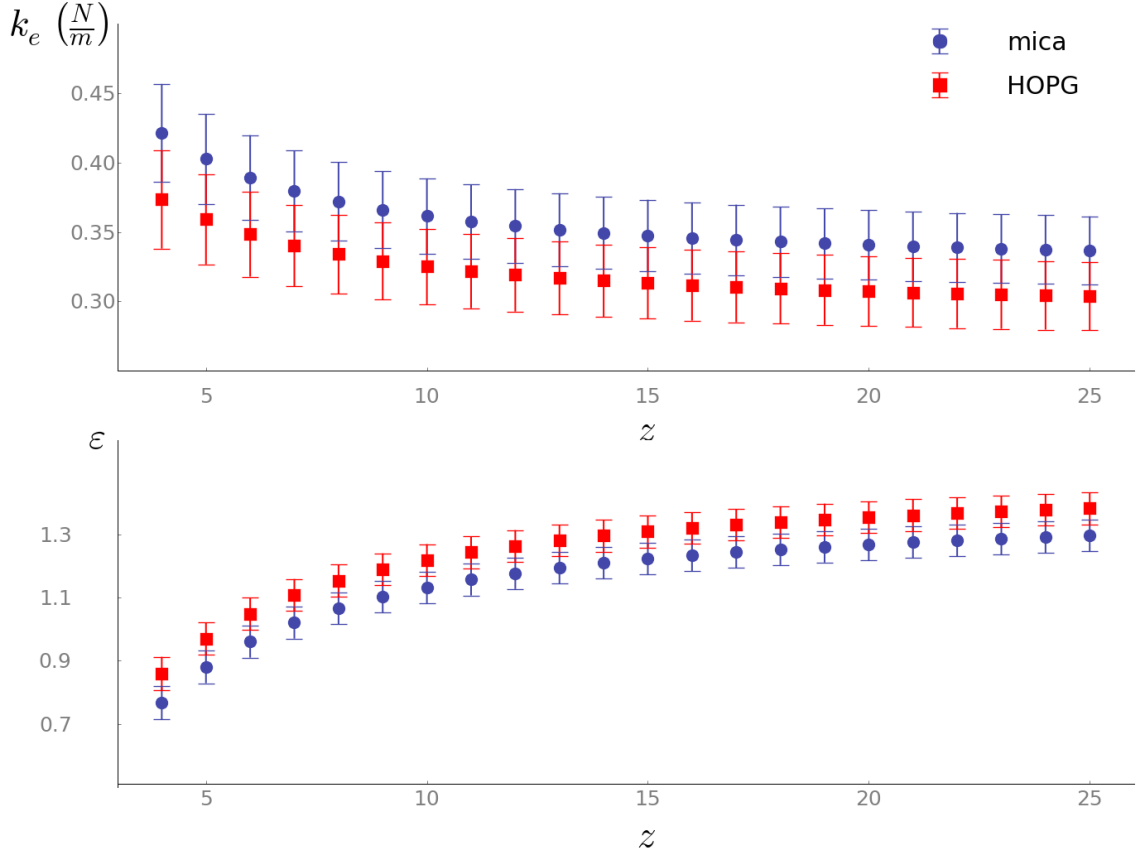
Figure 1.17 – Fit parameters as a function of z. Top: dependence of the fitted value for the spring constant k_e on the charge z. Bottom: dependence of ε on z. The blue circles and red squares correspond respectively to the data obtained on mica and HOPG. In both cases, the values of the parameters change very slightly when varying z, reaching in practice a plateau for $z > 10$. Reprinted from [67].

the experimental data by means of equation (1.51) for several values of z, and subsequently computed the mean spacing Δx. In fig.1.16 we show the results obtained for both mica (blue circles) and HOPG (red squares) for values of z in the range $4 \le z \le 12$. The horizontal line corresponds to an ordinate equal to l_B, so that the most physically suitable value of z is the one leading to a Δx as much close as possible to this line, which from the graph we deduce to be $z = 6$.

In fig.1.17 we report the values of the fitting parameters k_e and ε which would be obtained by considering different values of z. As we can see from the graph, these values are virtually independent of the actual value of z, that as a consequence does not affect our conclusions. Mathematically speaking, this can be explained by looking carefully at the formulas describing the electrostatics. Indeed, the value of z only appears in the equation defining the potential ψ_0, and thus ultimately appears in the formula of the pitch by means of the interaction
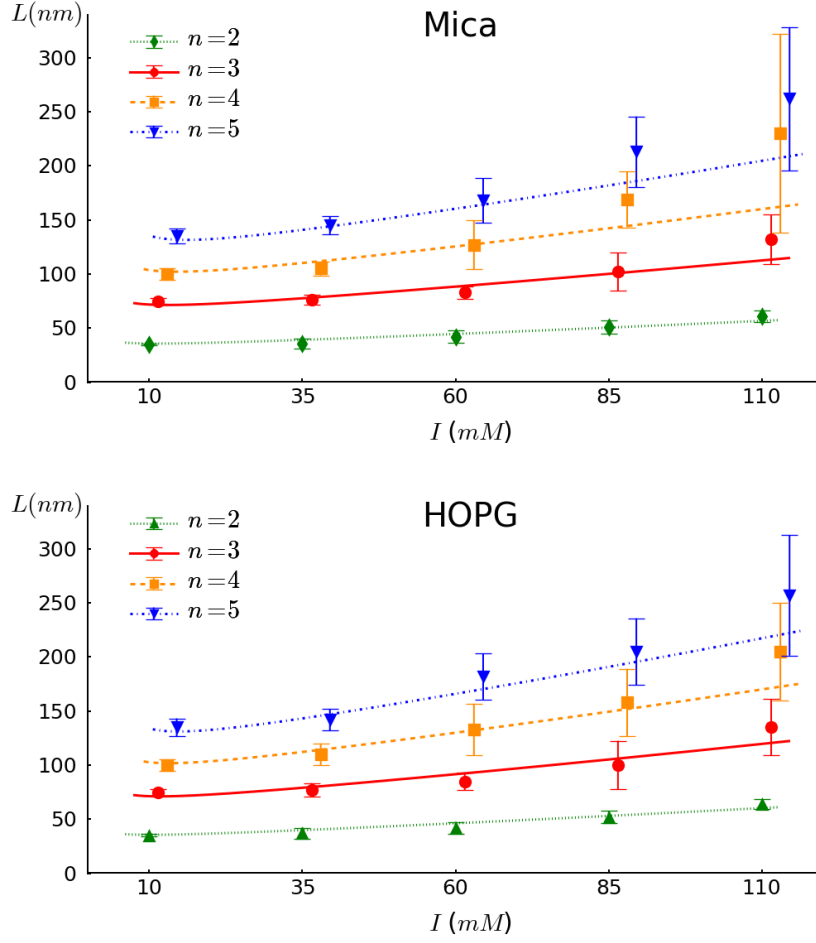
Figure 1.18 – Comparison between the experimental values of the halfpitches and theory after fit for fibrils adsorbed on mica and HOPG surfaces. In each case, all the 20 experimental points were fitted simultaneously by tuning the parameters b and ε. Reprinted from [67].

parameter $Z(I)$. More in detail, the potential is an increasing function of z, since it is basically its hyperbolic arcsinus. The interaction parameter Z depends on ψ_0 by means of a hyperbolic tangent, so that it is a monotonically increasing function which eventually reaches a plateau. Thus, there is a saturation effect, which as shown in fig.1.17 is already reached for relatively modest values of z.

The best fits obtained for $z = 6$ lead to the curves reported in fig.1.18. The final values of the parameters are $k_e = (0.39 \pm 0.03)$ N/m, $\varepsilon = 0.96 \pm 0.05$ for mica, and $k_e = (0.35 \pm 0.02)$ N/m, $\varepsilon = 1.05 \pm 0.04$ in the case of HOPG. From the fitted value of k_e it is possible to give an estimate of the torsional elastic properties of fibrils. In particular, apart from numerical prefactors which depend on n and ϵ (both of order unity), the shear modulus G is given by $G \sim k_e/b \sim 10^{-1}$ GPa, in agreement with previously reported experimental results [50]. Moreover, the fact that the best-fit values of the $k_e : \varepsilon$ pair differ by 10% or less between the

two substrates, which have very different surface energies, conclusively demonstrates that surface adsorption effects do not strongly affect the measured values of fibrils periodicity. These effects are relevant only at large values of ionic strength. Indeed, as already noticed in [44], for large I and n the fibrils appear as smashed on the surface due to adhesion forces. Though being always within the error bars, in this region the theoretical prediction seems to systematically underestimate the values of the halfpitch. In the next section we consider a refined model where the interaction with the surface is explicitly taken into account. Within this model, we show for HOPG, where no additional eletrostatic interactions play a role, that this allows to further improve the accuracy of the model in the regime of high pitches (large I and n).

Before delving into the modelling of fibril-substrate interactions, we note that the fitting parameters within the explicit model can be used to perform a sanity check of our calculations. In this regard, we now address the validity of some of the approximations done in order to arrive at equation (1.51) for the halfpitch. Particularly, based on a self-consistency argument, in section 1.3.2 we recognized a leading term in the exact perturbative result and discarded the remaining ones, finally obtaining equation (1.23). In a similar fashion, in section 1.4.1 we neglected one term in order to obtain equation (1.50). To further check the validity of such approximations, we now compute the discarded quantities within the explicit model and compare them with the leading term. In particular, by making use of the expression under square root in equation (1.50), we define the following ratios:

$$
r_1(n, I) \equiv \frac{\frac{1}{6}}{\frac{3n^2 - 7}{10} \frac{k_e e^{kb\varepsilon}}{kZ(1+\varepsilon)}} = \frac{5}{3} \frac{1}{3n^2 - 7} \frac{kZ(1+\varepsilon)}{k_e e^{kb\varepsilon}} \,,
$$

$$
r_2(I) \equiv \frac{\frac{3n^2 - 7}{40(1+\varepsilon)^2}}{\frac{3n^2 - 7}{10} \frac{k_e e^{kb\varepsilon}}{kZ(1+\varepsilon)}} = \frac{1}{4} \frac{1}{1+\varepsilon} \frac{kZ}{k_e e^{kb\varepsilon}} \,, \tag{1.52}
$$

$$
r_3(I) \equiv \frac{\frac{3n^2 - 7}{40} \frac{kb}{1+\varepsilon}}{\frac{3n^2 - 7}{10} \frac{k_e e^{kb\varepsilon}}{kZ(1+\varepsilon)}} = \frac{1}{4} kb \frac{kZ}{k_e e^{kb\varepsilon}} \,.
$$

We note that, for each value of I, the ratio r_1 as a function of n takes its largest value for $n = 2$, so that we will limit our analysis to this case. By substituting the values for k_e, ε obtained from the fit, we find the curves shown in fig.1.19. From the plot, both for mica and HOPG the largest value of such ratios is around $2 \cdot 10^{-2}$, thus providing an *a posteriori* validation of our assumptions. Finally, as a further check we fitted the data by using the formula for the pitch we would have obtained by not discarding any of the terms above. The results obtained are $k_e = (0.38 \pm 0.03)$ N/m, $\varepsilon = 0.97 \pm 0.05$ for mica, and $k_e = (0.34 \pm 0.03)$ N/m, $\varepsilon = 1.05 \pm 0.05$ for HOPG, which are practically the same as the ones obtained with the approximated formula (compare with the numerical values provided above).
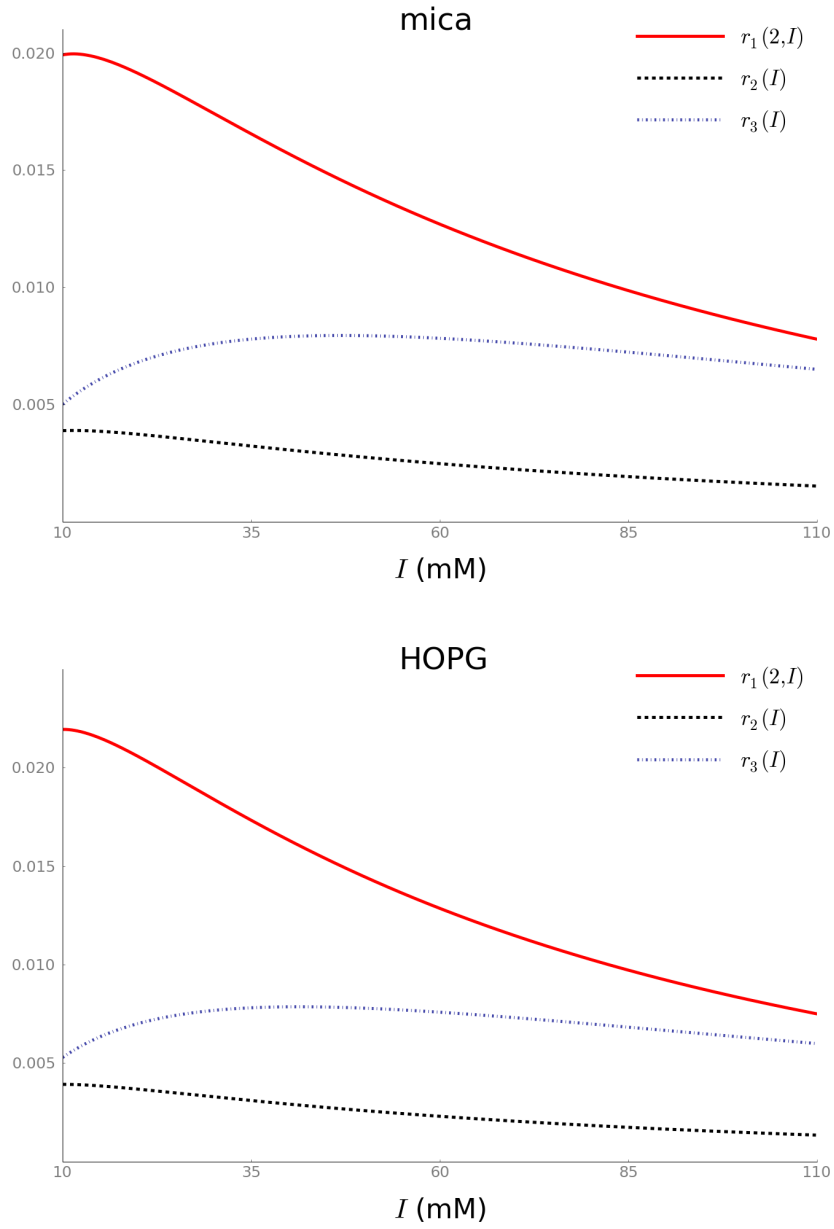
Figure 1.19 – Check on the validity of the approximations done. Top: plot of the ratios defined in equation (1.52) after the fitting values of k_e, ε found for mica have been substituted. Bottom: plot of the ratios in the case of HOPG fitting parameters. Reprinted from [67].

1.4.3 Toy model of fibril-substrate interactions.

In this section, we introduce a model that mimicks the effect of the interaction of the system with the substrate on the pitch of the fibrils. This interaction is introduced in the simplest way possible, namely through a harmonic potential between beads and substrate, here represented as a homogeneous plane. This approach is naturally too simplistic to capture the details of the complex interaction between fibrils and substrates (in particular in the case of mica, where this interaction is of electrostatic nature), but gives useful insights on how the presence of a substrate influences the bulk values of the pitch.

In what follows, we will assume each bead to interact independently with the surface via a potential dependent only on the distance substrate-bead. By introducing this interaction, the system is no more symmetric with respect to the longitudinal axis of the fibril. From a local perspective, the net interaction between two consecutive layers is now expected to depend not only on the twist angle ϕ, but also on the orientation of the layers with respect to the surface. Nevertheless, the system still presents a symmetry with respect to the configuration of a vertically adsorbed ribbon, so that from now on we will restrict our analysis to the portion of the pitch between a horizontally and a vertically adsorbed fibril, *i.e.* to half of the halfpitch L.

Let us assume that each bead interacts harmonically with the surface, *i.e.* it is subject to a potential $\bar{k}_e h^2/2$, where h is the closest approach distance between the bead and the surface (from now on we will refer to h as *height* of the bead). Assuming that each layer has exactly one[3] contact point with the substrate (apart from the horizontally adsorbed configuration), if we label the beads belonging to it as $0,\ldots,n-1$ the height h_i of the ith bead is given by $h_i = i b \sin\Theta$, where $\Theta \in [0,\pi/2]$ is the angle formed by the layer and the plane. Now, let us consider two consecutive layers twisted by an angle ϕ and let θ be the angle formed by the fibril and the surface at the mid-point between them. Therefore, the two layers are oriented with angles $\theta - \phi/2$ and $\theta + \phi/2$ with respect to the substrate. The total interaction between layers and plane thus reads

$$\frac{1}{2}\bar{k}_e \sum_{i=0}^{n-1} (ib)^2 \left(\sin^2(\theta - \phi/2) + \sin^2(\theta + \phi/2)\right). \tag{1.53}$$

By making use of the formulas

$$\sin^2 x = \frac{1 - \cos 2x}{2}, \tag{1.54}$$

$$\cos x + \cos y = 2\cos\frac{x+y}{2}\cos\frac{x-y}{2}, \tag{1.55}$$

---

[3]From a geometrical perspective, such an assumption is incompatible with a perfectly straight fibril axis, which now will cover a sinusoidal trajectory along one period. However, given the values of n, b and L the length of such a sine is practically that of the corresponding end-to-end straight line, so that to our purposes the axis can be still considered as perfectly straight.

$$\sum_{i=0}^{n-1} i^2 = \frac{1}{6} n(n-1)(2n-1) \tag{1.56}$$

the previous result can be rewritten as

$$\frac{1}{12} n(n-1)(2n-1)\bar{k}_e b^2 \frac{2 - \cos(2\theta + \phi) - \cos(2\theta - \phi)}{2} =$$

$$= \frac{1}{12} n(n-1)(2n-1)\bar{k}_e b^2 \frac{2 - 2\cos 2\theta \cos\phi}{2}. \tag{1.57}$$

By Taylor-expanding $\cos\phi$ and neglecting the additive ϕ-independent terms, we finally obtain

$$\frac{1}{12} n(n-1)(2n-1)\bar{k}_e b^2 \cos 2\theta \left(\frac{1}{2}\phi^2 - \frac{1}{24}\phi^4\right). \tag{1.58}$$

The potential $V_I^r(\phi)$ in equation (1.18) can be rewritten within the explicit model as (by retaining only the leading terms)

$$V_I^r = -n(n^2 - 1)\frac{kb^2 Z e^{-kb\varepsilon}}{96(1+\varepsilon)}\phi^2 + n(n^2 - 1)(3n^2 - 7)\frac{k_e b^2}{1920(1+\varepsilon)^2}\phi^4. \tag{1.59}$$

By defining $\rho_I \equiv kb^2 Z \exp(-kb\varepsilon)/96(1+\varepsilon)$ and $\eta \equiv k_e b^2/1920(1+\varepsilon)^2$, the previous formula can be rewritten as $V_I^r = -n(n^2 - 1)\rho_I \phi^2 + n(n^2 - 1)(3n^2 - 7)\eta\phi^4$. Note that the unperturbed equilibrium angle ϕ^* can be written in terms of these new parameters as $\phi^* = \sqrt{\rho_I/2\eta(3n^2 - 7)}$. Moreover, as such formula for ϕ^* suggests, ρ_I is always sensibly smaller than η. Finally, by further defining $\lambda_n \equiv 1/288(2n-1)/(n+1)\bar{k}_e b^2$, the total energy U_θ of the two-layers system with surface effects can be obtained as the sum of (1.58) and (1.59) and reads

$$\frac{U_\theta}{n(n^2 - 1)} = \left(-\rho_I + 12\lambda_n \cos 2\theta\right)\phi^2 + \left[\eta(3n^2 - 7) - \lambda_n \cos 2\theta\right]\phi^4. \tag{1.60}$$

By minimizing the previous formula with respect to ϕ and rearranging, we find, apart from the trivial solution $\phi = 0$,

$$\bar{\phi}^2 = \frac{1 - 12\frac{\lambda_n}{\rho_I}\cos 2\theta}{1 - \frac{\lambda_n}{\eta(3n^2 - 7)}\cos 2\theta}\phi^{*2}, \tag{1.61}$$

that is, remembering that $\phi^{*2} = \rho_I/2\eta(3n^2 - 7)$,

$$\bar{\phi}^2 = \frac{1 - 12\frac{\lambda_n}{\rho_I}\cos 2\theta}{1 - \frac{2\lambda_n\phi^{*2}}{\rho_I}\cos 2\theta}\phi^{*2} \simeq \left(1 - 12\frac{\lambda_n}{\rho_I}\cos 2\theta\right)\phi^{*2}. \tag{1.62}$$

By extracting the square root of both sides in the previous equation, we are thus finally able to write the formula for $\bar{\phi}$:

$$\bar{\phi} = \phi^* \sqrt{1 - 12 \frac{\lambda_n}{\rho_I} \cos 2\theta}. \tag{1.63}$$

From equation (1.63) we can already draw an important conclusion: according to our model, for $\theta < \pi/4$ the perturbed equilibrium angle $\bar{\phi}$ is smaller than ϕ^*, *i.e.* the fibril twists more slowly with respect to the unperturbed case. In contrast, $\theta > \pi/4 \Rightarrow \bar{\phi} > \phi^*$, *i.e.* the twist is fastened. In other words, the qualitative picture emerging from this result is that of a fibril which twists slowly while being close to a horizontally adsorbed configuration ($\theta < \pi/4$), but increases its $\bar{\phi}$ once it gets close to the vertically adsorbed one ($\theta > \pi/4$), in agreement with the empirical observations in ref. [44].

Starting from equation (1.63), in order to compute the pitch we have to find a set of angles $\bar{\phi}_j, j = 0, \dots, \bar{N} - 1$ such that:

1. $\sum_{j=0}^{\bar{N}-1} \bar{\phi}_j = \pi/2$;

2. $\bar{\phi}_j = \bar{\phi}(\theta_j)$;

3. $\theta_0 = 0$ and $\theta_j = \theta_{j-1} + \phi_{j-1}/2 + \phi_j/2$ for $j > 0$.

Notice that the number $\bar{N}$ is not fixed *a priori*. Quite the contrary, the final aim of this calculation is the very computation of $\bar{N}$, since the half-pitch will then simply be $\bar{L} = 2\bar{N}b(1+\epsilon)$, where the factor of two comes from the fact that in order to get a halfpitch we have to consider twice the portion of fibril under inspection. Therefore, in the very end the effect of the presence of the surface on the pitch can be understood by comparing $\bar{N}$ with its analogue in the unperturbed case $N^* \equiv \pi/(2\phi^*)$.

By looking at points 2. and 3. above, we note that once θ_{j-1} is known, in order to find θ_j we need to solve a transcendental equation, so that the problem cannot be solved analitically. Nevertheless, we can still get some qualitative insight into the behavior of the pitch with some simple arguments.

First of all, by defining $a \equiv 12 \frac{\lambda_n}{\rho_I}$, we notice that the average value of $\bar{\phi}(\theta)$ over the interval $[0, \pi/2]$ is equal to

$$\langle \bar{\phi} \rangle_{\mathrm{cont}} = \phi^* \frac{2}{\pi} \int_0^{\frac{\pi}{2}} d\theta \sqrt{1 - a \cos 2\theta} = \phi^* \frac{2}{\pi} \int_0^{\frac{\pi}{2}} d\theta \sqrt{1 + a - 2a \cos^2 \theta} =$$

$$= \phi^* \frac{2}{\pi} \int_0^{\frac{\pi}{2}} d\varphi \sqrt{1 + a - 2a \sin^2 \varphi} = \phi^* \frac{2\sqrt{1+a}}{\pi} E\left(\sqrt{\frac{2a}{1+a}}\right) \equiv \phi^* f(a), \tag{1.64}$$

where we have performed the change of dummy variable $\varphi \equiv \pi/2 - \theta$,

$$E(x) \equiv \int_0^{\pi/2} d\varphi \sqrt{1 - x^2 \sin^2 \varphi} \tag{1.65}$$

is the complete elliptic integral of the second kind and we have defined the function

$$f(a) \equiv \frac{2\sqrt{1+a}}{\pi} E\left(\sqrt{\frac{2a}{1+a}}\right). \tag{1.66}$$

The function $f(a)$ monotonically decreases in the range $[0,1]$ (the only interval allowed for the surface effects not to disrupt the fibril, see below), thus it attains its maximum value for $a = 0$, which is equal to $f(0) = 1$. Therefore, the largest value attained by $\langle\bar\phi\rangle_{\text{cont}}$ is in absence of surface effects, where as expected $\langle\bar\phi\rangle_{\text{cont}} = \phi^*$. Now, assume to have found a set of $\bar\phi_j$ fulfilling the requirements summarized in points 1,2,3 above. By definition, their average $\langle\bar\phi\rangle$ is such that $\bar N \langle\bar\phi\rangle = \pi/2$. Since $\bar\phi_j \ll 1$, at a first approximation we can assume that the average of such a discrete set is equal to its continuous version, i.e. $\langle\bar\phi\rangle \simeq \langle\bar\phi\rangle_{\text{cont}}$. At this point, the computation of $\bar N$ becomes straightforward:

$$\bar N \langle\bar\phi\rangle_{\text{cont}} = \pi/2 = N^* \phi^* \Rightarrow \bar N = N^* \frac{\phi^*}{\langle\bar\phi\rangle_{\text{cont}}} = \frac{N^*}{f(a)}. \tag{1.67}$$

Finally, remembering that $\bar L = 2\bar N b(1+\varepsilon)$ and $L = 2N^* b(1+\varepsilon)$, we obtain for the half-pitch

$$\bar L(a) = \frac{L}{f(a)}. \tag{1.68}$$

Since $f(a) < 1$ for $a \neq 0$, from Equation (1.68) we argue that as a net effect the surface effects result into an increase of the half-pitch. Moreover, because of the monotonicity of $f(a)$ such an increase is more marked for larger values of a. Remembering that $a = 12\lambda_n/\rho_I$, and substituting λ_n and ρ_I by means of their definitions, we obtain the explicit dependence of a on n, I:

$$a(n, I) = 4\frac{2n-1}{n+1} \frac{\bar k_e(1+\varepsilon)e^{k(I)b\varepsilon}}{k(I)Z(I)}. \tag{1.69}$$

The formula above is an increasing function of both n and I, so that we conclude that within this simple model the surface effects on the pitch are to be stronger for wider fibrils (i.e. larger n) and more salty solutions, which can qualitatively explain the discrepancy between theory and experiments found in the framework of our model.

Though simplistic, the present approach is thus able to take into account the slight under-estimation of the pitch in the unperturbed model, and qualitatively predicts the observed dependence of such differences upon the ionic strength. Interestingly, a fit of the HOPG data performed by tuning k_e, ε and $\bar k_e$ and considering the proper computations made by following points 1, 2, 3 above leads to a better agreement between theoretical predictions and

experimental data, as we show in fig.1.20 (compare fig.1.18 bottom). The final values for the parameters are $\varepsilon = 0.95$, $k_e = 0.38$ N/m, $\bar{k}_e = 3.7 \cdot 10^{-4}$ N/m. A comparison with the values obtained in the unperturbed case shows that the parameters are not strongly affected by the introduction of the new interaction. In spite of the good agreement between theory and data, we want to stress that, due to the strongly oversimplified picture introduced in the present treatment, the value of the fitting parameter $\bar{k}_e$ is not quantitatively reliable, but serves more as a rough measure of the expected strength of fibril-surface interactions.
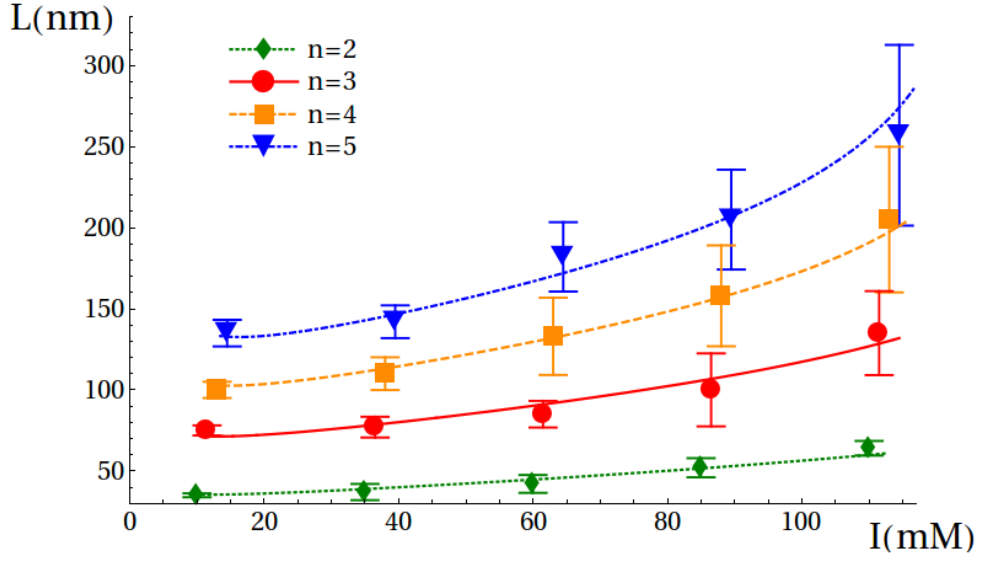
Figure 1.20 – Half-pitches of fibrils adsorbed on HOPG as a function of the ionic strength I for $n = 2, 3, 4, 5$ and corresponding fitted prediction in the presence of fibril-substrate interactions. Reprinted from [67].

As a final remark, it is interesting to observe that the present model predicts the existence of a critical value of ionic strength I beyond which the interaction with the substrate leads to the disruption of the fibril. Indeed, let us consider again equation (1.63): as long as $a = 12\lambda_n / \rho_I < 1$ the argument of the square root will be positive for any values of θ, so that all our considerations above apply. However, once a becomes larger than the critical value $a_c = 1$ (corresponding to a certain value I_c of the ionic strength), for small enough θ we get $1 - a\cos 2\theta < 0$, so that the only extremal point of the free energy will be $\phi = 0$, which switches from being a maximum to a minimum of the system[4]. In other words, in this situation the

---

[4]This is a well-known behavior in statistical physics, where a Landau expansion of the free energy like the one we did in equation (1.60) can be seen as the behavior of the order parameter - ϕ in our case - in the vicinity of a

interaction fibril-substrate is too strong, and the minimum of the free energy of the system is obtained by completely relaxing the fibril, *i.e.* by disrupting the twist. The stronger the interaction, the larger the range of values of θ resulting in the disruption of the fibril.

1.5 Conclusions

We have shown here that the trade-off between the competing energetic contributions of electrostatics and elasticity governs the mesosopic architecture of amyloid fibrils. Our model quantitatively predicts the dependence of the pitch of the twist of the fibers upon the number of protofilaments and on the physical parameters, and in doing so demonstrates the existence of hitherto unknown universal, parameter-free features, bringing an unprecedented new description on the physical mechanisms controlling amyloid fibrils self-assembly. More generally, our results should apply whenever two different competing interactions, not necessarily of elastic and electrostatic nature, destabilize the flat architecture of the ribbon, so to give rise to twisted fibrils. From an even broader perspective, our findings suggest that there might be a number of yet undiscovered universal laws governing several mesoscale features of protein fibrillar aggregates.

---

continuous phase transition.

2 Coarse-grained modelling of Hsp70

2.1 A primer on Hsp70

Molecular chaperones are very versatile proteins that assist other proteins in various cellular processes [78]. Among them, the 70-kDa heat-shock protein (Hsp70) is probably the most eclectic one, being involved in a wide variety of cellular processes both in prokaryotes and eukaryotes [78–80]. Historically, Hsp70s were given this name since they were overexpressed when putting bacterial cells in very harsh conditions, such as high-temperature baths [81]. Since then, Hsp70 molecules were systematically found to assist other proteins in a plethora of cellular processes [79]. For example, Hsp70s prevent the aggregation of proteins [82] and supervise the disaggregation of stable protein aggregates [83]; they assist proteins in *de novo* folding, particularly when the latter have large size [84] or in the case of mutants [85]. Hsp70 chaperones also interact with many signal-transduction proteins [86], intervene in protein-complexes remodeling such as the disassembly of clathrin cages [87] and drive posttranslational import of cytoplasmic proteins into eukaryotic organelles [88–90]. Moreover, it has been recently outlined that they may also be involved in preventing translation stalling in ribosomes [91, 92].

From a structural point of view, Hsp70s are among the most conserved proteins throughout evolution [95]. Because of this, results found on a particular member of the Hsp70 family are usually extended to all of them [78]. As a consequence, the scientific literature has become more and more focused on DnaK, the Hsp70 chaperone of *E. coli*, which has thus been appointed as the model case for this class of molecules [78]. DnaK is an ATP-consuming protein made of 600 amino acids [78]. Its structure rougly consists of two large domains connected by a small flexible linker (see fig.2.1). Specifically, the N-terminal Nucleotide Binding Domain (NBD) is the ATPase unit of the chaperone, while the Substrate Binding Domain (SBD) directly interacts with specific sites on the target protein. These binding sites are frequently found in protein sequences [96], so that multiple chaperones are likely to bind the same substrate.

According to the current view of the Hsp70 biochemical cycle [98], DnaK can be typically found
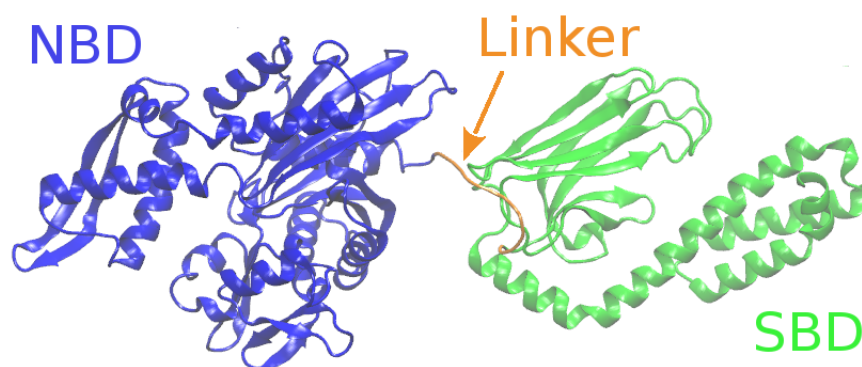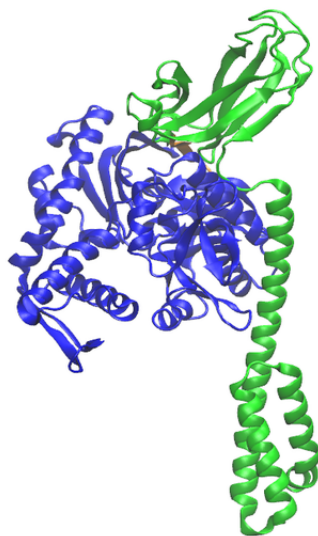
Figure 2.1 – Cartoons of ADP-, substrate-bound *E. coli* DnaK (top panel, PDB: 2KHO [93]) and ATP-bound chaperone in the absence of the substrate (bottom panel, PDB: 4B9Q [94]).

in two distinct states, depending on whether NBD is hosting ATP or ADP. As shown in table 2.1, the two states have very different affinities for the substrate, with the ATP-bound state

	K_D	k_{on} (M^{-1}s^{-1})	k_{off} (s^{-1})
ADP	63 nM	9.3 x 10^3	0.004
ATP	2.2 μM	4.5 x 10^5	1.8

Table 2.1 – Dissociation constant and kinetic rates for substrate binding by ATP- and ADP-bound Hsp70. Data taken from [78, 97].

showing a dissociation constant equal to roughly 35 times that of its ADP counterpart [97]. The kinetic constants show even larger differences. The ATP-bound state has a large binding rate when compared to the ADP-bound case, while the latter has a much lower value of k_{off}. It has been proposed that Hsp70 exploits this large difference in time scales by means of a mechanism called ultra-affinity [99]: the energy released by ATP hydrolysis is used to maintain the chaperone in a nonequilibrium steady-state cycle which takes advantage of the k_{on} of the ATP-bound state and the k_{off} of the ADP-bound state. In doing so, DnaK shows an effective dissociation constant orders of magnitude lower than the K_d of both the two equilibrium states [99].

The ultra-affinity mechanism is intimately related to the experimentally-observed increase in ATP hydrolysis upon substrate binding [99]. Since the two events are spatially localized in regions having a mutual distance of more than 50 Å [78] (the former takes place in the NBD, while the latter involves the SBD), it does not come as a surprise that the presence of allosteric mechanisms has been observed [78]. In this regard, many experiments have been performed using several approaches, including (but not limited to) NMR spectroscopy [100, 101] and fluorescence techniques [102, 103]. To summarize, substrate binding catalyzes ATP hydrolysis, which in turn stabilizes the former. Moreover, ATP binding in the NBD induces substrate release [78]. From this picture, one concludes that the key conformational states are the ADP-, substrate-bound DnaK and the ATP-bound state in absence of the substrate. The structures of DnaK in these two states have been recently unveiled [93, 94] and are reported as ribbon cartoons in fig.2.1. In the ADP-bound state (fig.2.1 top), the two domains can move independently, apart from the constraint imposed by the presence of the flexible linker [78]. The bound substrate (not shown in fig.2.1) is localized between the two β-sheets of the SBD, while the C-terminal α-helices (called LID) shield it from interactions with the environment [78]. In the ATP-bound state, the surfaces of the NBD and the SBD are constantly in contact, while the linker is stably docked within the core of the NBD, where it interacts with its subdomains. As a result, the domains are strongly constrained to each other, and the chaperone behaves approximately as a rigid body [78]. Moreover, the LID and the β-sheets subdomain of the SBD are well apart from each other, so that the binding region of the latter is exposed to the solvent. These observations provide a molecular explanation to the different scales in the kinetic constants reported in table 2.1: on the one hand, in the ATP-bound state the solvent exposure of the binding site eases the interaction with the substrate, resulting in fast kinetics; on the other hand, in the ADP-bound state the binding region is hindered from the solvent, and both substrate release and binding are rather unlikely events. In a more

pictorial way, the chaperone "bites" the substrate, and the ATP-ADP exchange mechanism acts as a switch between the open and closed states.

Finally, it is worth mentioning that Hsp70 is assisted by other two classes of co-chaperones, namely the J-domain proteins (JDP) and the Nucleotide Excange Factors (NEF), which in *E. Coli* are named DnaJ and GrpE. A JDP, also known as Hsp40 [78], is characterized by the presence of a J-Domain, which interacts with Hsp70, and other domains with similar substrate preferences to DnaK [78, 104, 105]. Moreover, it has been shown that it also catalyzes the hydrolysis of the ATP molecule in the NBD [104]. As a net result, JDPs act therefore as recruiters of Hsp70, further increasing their binding rate. Analogously, the NEFs affect the k_{off} of the chaperones by inducing the release of ADP from the NBD, which is promptly substituted by an ATP molecule thanks to the high affinity of NBD for nucleotides and to the relative large *in vivo* concentration of ATPs with respect to ADP molecules. Interestingly, thanks to ist elongated shape NBD-bound GrpE also competes with the substrate to establish an interaction with the SBD, thus facilitating its release [106].

The omnipresence of Hsp70 in a huge variety of cellular processes suggests that this chaperone works by means of a general mechanism. In this regard, based on excluded-volume effects, the *entropic pulling* mechanism proposed in [107] provides for a simple way to explain the role played by Hsp70 in this myriad of processes [80]. In this chapter, we focus on a specific case, namely the posttranslational import of cytoplasmic proteins into organelles. Starting from the entropic-pulling mechanism, we compute by means of MD simulations the effective free-energy profile for substrate translocation induced by the presence of Hsp70. This free energy is then used to quantitatively characterize the import process, thus outlining the key role of Hsp70 in protein import into mitochondria.

2.2 Hsp70 as translocase

2.2.1 The entropic-pulling mechanism

Though mitochondria have their own DNA, the majority of mitochondrial proteins are actually encoded in nuclear DNA, synthesised in the cytosol and only post-translationally imported into the organelles [88]. Protein import takes place through a proteinaceous pore that spans the two mitochondrial membranes by way of the outer (TOM) and inner (TIM) membrane pore complexes [88]. According to the current view, an ATP-consuming import motor located into the mitochondrial matrix drives the inward translocation of nuclear-encoded proteins. Mitochondrial Hsp70 (mtHsp70) is the central element of this motor: it is recruited by the TIM complex on the matrix side through interactions with the TIM44 protein, which is part of the pore, and with the pore-associated PAM16/18 proteins. The latter contain a J domain, therefore they act as recruiters and increase by orders of magnitude the affinity of chaperones for substrates. The ATP-driven binding of Hsp70s to incoming proteins ultimately drives their translocation.

The precise mechanism by which Hsp70 exerts its pulling action has been debated in the literature and several models have been proposed [88, 107, 108]. The *Brownian ratchet* [88] assumes that, thanks to the chaperone large size, Hsp70 binding prevents the retrotranslocation of the substrate into the pore, thus biasing the random fluctuations toward the matrix. Alternatively, according to the *power stroke* [108] the chaperone actively pulls the incoming protein by using TIM as a fulcrum. Later, according to the *entropic pulling* model [107], it was shown that an active force naturally emerges from a realistic physical description of the Brownian ratchet, thus reconciling the two views [80]. Indeed, the excluded volume of the chaperone, besides preventing retrotranslocation, reduces the conformational space available to the incoming protein, thus decreasing its entropy. This reduction depends on the length of the imported fragment of the substrate, therefore resulting in a free-energy gradient which favors the import.
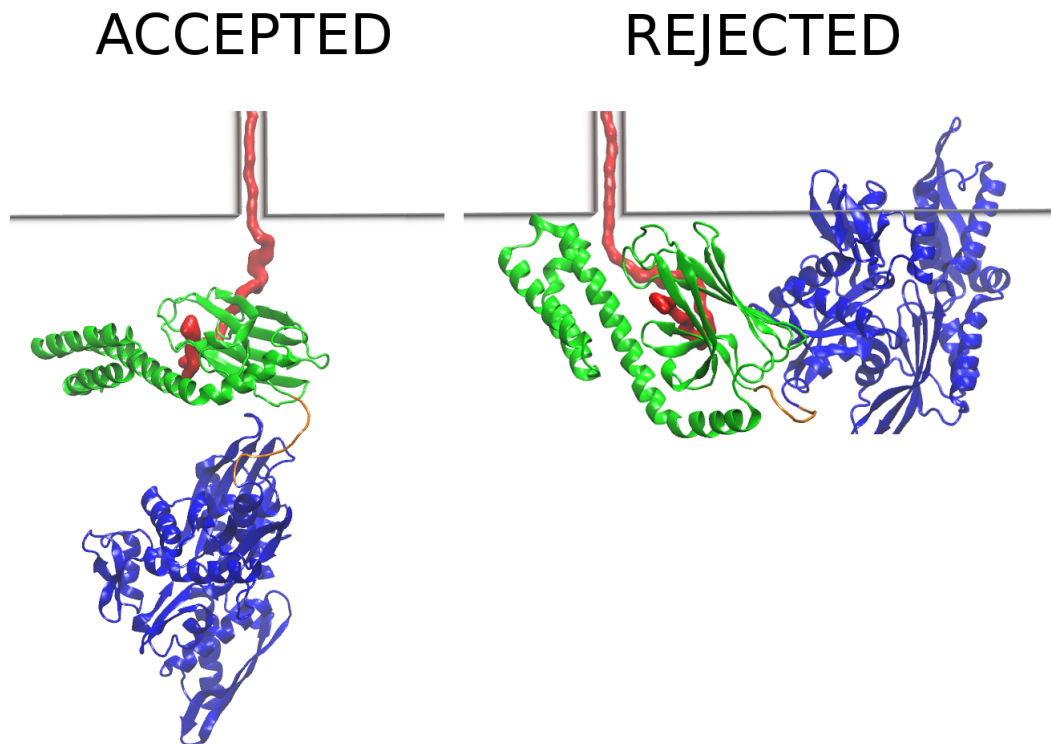
Figure 2.2 – Effect of the excluded volume of Hsp70 on the import process of a substrate. The substrate-bound chaperone is depicted in its ADP-bound state with the same color code as in fig.2.1, while the substrate is represented as a structureless red tube. The inner mitochondrial membrane is described as a horizontal flat plane. In the absence of chaperone, both the substrate conformations reported are sterically allowed. In contrast, with a bound Hsp70 the right one would result in a clash between Hsp70 and membrane.

This mechanism is best explained with the aid of fig.2.2, where we sketch a simplified description of the system under study. The chaperone is represented at the secondary-structure level in its ADP-bound state, with the same color code as in fig.2.1. The substrate, depicted

as un unstructured red tube, is being imported through the mitochondrial pore, represented here as a straight tunnel carved into the inner mitochondrial membrane, described as a flat plane. The effect of the size of the chaperone is two-fold. On the one hand, bound Hsp70 prevents the retrotranslocation of the substrate beyond its binding point (Brownian ratchet model [88]). On the other hand, the size of the chaperone leads also to a reduced number of sampled conformations (entropic pulling [107]), an effect not accounted for by the Brownian ratchet as it was originally conceived, but nonetheless intimately related to the same physical mechanism. For example, in the absence of Hsp70 the two substrate conformations shown in fig.2.2 are both sterically allowed. However, upon chaperone binding the conformation on the right would result into an overlap between the membrane and Hsp70, and it is therefore never sampled by the substrate when the chaperone is present.

To get a more quantitative insight, we now report the analytical computations performed in [107], where the authors consider a strongly simplified version of the system. Specifically, the chaperone is represented as a sphere with radius $R_{\text{chap}} = 20$ Å, while the substrate is modelled as a Freely-Jointed Chain (FJC) starting from the exit of the pore. The latter provides the simplest representation of a polymer, which is described as a random walk with step length l [27]. Though simplistic, this approach can be successfully used whenever one can neglect the interaction between monomers separated by large contour lengths (due *e.g.* to excluded volume, electrostatics or hydrophobicity). Indeed, in this case only local interactions are important, and for long enough chains one can remap the statistical properties of more complicated polymers onto equivalent FJCs with a renormalized step size, called Kuhn length [27]. In the present case, a Kuhn length equal to $l = 12$ Å is used [109]. In the case of an unrestricted FJC made of n monomers (that is, $n-1$ bonds), the probability distribution of the end-to-end vector $\boldsymbol{R_e}$ is a three-dimensional Gaussian function [16]:

$$P_n^G(\boldsymbol{R_e}) = \left(\frac{3}{2\pi(n-1)l^2}\right)^{\frac{3}{2}} e^{-\frac{3\boldsymbol{R_e^2}}{2(n-1)l^2}} \tag{2.1}$$

However, this function is not suitable to describe the present system, since both Hsp70 and the membrane have to be considered. In this regard, the chaperone is centered on the last monomer of the FJC and no Hsp70-substrate interactions are considered. The mitochondrial membrane is represented as a flat wall corresponding to the xy plane, which is impenetrable to both the substrate and the chaperone. By considering a reference frame with the z axis perpendicular to the membrane and directed towards the mitochondrial matrix, when the chaperone is absent the conformational space encompassed by the substrate can be obtained by rejecting all the conformations having at least one monomer with $z < 0$ (because of the presence of the wall). In this case, it can be shown that the following modification to equation (2.1) holds [16]:

$$P_n(\boldsymbol{R_e}) \propto \left(\frac{3}{2\pi(n-1)l^2}\right)^{\frac{3}{2}} \frac{6z}{(n-1)l^2} e^{-\frac{3\boldsymbol{R_e^2}}{2(n-1)l^2}}, \tag{2.2}$$

where the proportionality constant can be fixed by imposing normalization. Integrating over

x and y, we can thus retrieve the distribution probability for z:

$$P_n(z) = \frac{1}{Z(n)} \left(\frac{3}{2\pi(n-1)l^2} \right)^{\frac{1}{2}} \frac{6z}{(n-1)l^2} e^{-\frac{3z^2}{2(n-1)l^2}} \,, \tag{2.3}$$

where $Z(n)$ is the partition function of the system along the coordinate n in absence of the chaperone. An easy calculation shows that

$$Z(n) = \int_0^\infty \left(\frac{3}{2\pi(n-1)l^2} \right)^{\frac{1}{2}} \frac{6z}{(n-1)l^2} e^{-\frac{3z^2}{2(n-1)l^2}} = \left(\frac{6}{\pi(n-1)l^2} \right)^{\frac{1}{2}} \,. \tag{2.4}$$

Within the present assumptions, the effect of the chaperone can be straightforwardly included by furtherly restricting the conformational space. Specifically, all the conformations which are sterically allowed also in the presence of Hsp70 are those whose end-to-end vector satisfies $z \geq R_{\text{chap}}$. This restriction of the conformational space leads to a corresponding increase in entropy, which is in turn reflected in a quantitatively smaller partition function for the substrate in presence of Hsp70, $Z_{70}(n) < Z(n)$. A direct computation gives

$$Z_{70}(n) = \int_{R_{\text{chap}}}^\infty \left(\frac{3}{2\pi(n-1)l^2} \right)^{\frac{1}{2}} \frac{6z}{(n-1)l^2} e^{-\frac{3z^2}{2(n-1)l^2}} = \left(\frac{6}{\pi(n-1)l^2} \right)^{\frac{1}{2}} e^{-\frac{3R_{\text{chap}}^2}{2(n-1)l^2}} \,. \tag{2.5}$$

As a result, chaperone binding leads to a net increase in the conformational free energy of the substrate equal to

$$\Delta F_c(n) = -k_B T \ln \left(\frac{Z_{70}(n)}{Z(n)} \right), \tag{2.6}$$

which, after substitution of equation (2.4) and equation (2.5), gives

$$\Delta F_c(n) = \frac{3R_{\text{chap}}^2}{2(n-1)l^2} k_B T \,. \tag{2.7}$$

As one could have expected, the increase in free energy is largest for the lowest values of n. It is important to notice that it is not this increase *per se* that generates a net thermodynamic force facilitating protein import. Instead, the key point is that ΔF_c depends on n. In other words, the chaperone introduces a net free-energy *gradient*, which provides the driving force for the import.

It is worth noting that, in a general perspective, the presence of chaperones introduces an asymmetry between the interior and the exterior of the mitochondrial membrane. Traditionally, polymer translocation between asymmetric environments has been studied by introducing a chemical-potential difference μ between the two sides [110]. In the present case, however, since the typical size of a protein is around 300 amino acids [111] and since on average there is a binding site for Hsp70 every 35 residues [96], the number of chaperones that can bind simultaneously a substrate is on average about ten, making the mean-field picture not appropriate.

At the same time, the small number of bound Hsp70s makes the explicit contribution of the chaperones to the import process directly accessible. This level of detail enriches the picture of the system with non-trivial results such as the free-energy dependence on the number of imported residues, which is a natural consequence of the finite size of chaperones that in a mean-field picture should be artifically introduced in the model. Moreover, as stressed in equation (2.7), this size tunes the extent of the introduced free-energy gradient, which depends quadratically on R_{chap}. Again, this result has not a natural counterpart in the mean-field approach, where it should be artificially introduced. Equation (2.7) suggests that increasing the size of the chaperone would lead to more dramatic effects, speeding up the import process.

As a final remark, the generality of the proposed mechanism suggests that Hsp70 is not strictly needed for the import process, and it could be substituted by other large molecules. This is in line with the experimental results in [89], where it was shown that replacing Hsp70 with an antibody in ER import still leads to a successfull translocation.

2.2.2   Free energy computation from MD simulations

We now propose to compute the entropic-pulling free energy by considering a more detailed molecular model, where the conformational space of the substrate is explored by taking into account both the main structural features of Hsp70 and the local elastic properties of a protein [112]. To this aim, for each value of n we compute the effect of chaperone binding on the free energy of the system by means of coarse-grained Molecular Dynamics (MD) simulations. In the next section, this result is then used to devise a simplified yet quantitative analysis of the import, described as a one-dimensional diffusion process on the computed free-energy landscape.

Here, we compute the free energy difference $\Delta F_c(n)$ by estimating the ratio $Z_{70}(n)/Z(n)$ for n in $8 \leq n \leq 30$ with multiple coarse-grained MD simulations. The substrate was modeled as a n-residues flexible chain with the position of the nth residue constrained on the inner mithocondrial membrane, represented here as a flat wall acting only on the substrate residues. As a consequence, the system could sample configurations involving an overlap between the membrane and the chaperone (see right panel in fig.2.3). With this strategy, we could estimate the ratio $Z_{70}(n)/Z(n)$ as the fraction of time spent by the system in physically-acceptable, *i.e.* non-overlapping, configurations.

The MD simulations were performed coarse-graining both the substrate and the chaperone by considering one interaction site per residue centered on the C_α atom (compare fig.2.2 and fig.2.3). Residue-residue excluded-volume interactions were modeled with a repulsive Lennard-Jones potential with parameters $\sigma = 3.8\,\text{Å}$ and $\varepsilon = 3k_B T$. The substrate was modeled by using the local flexible potential introduced in [113]. Within this force field, the elastic properties of an unfolded protein are described by means of harmonic C_α-C_α bonds and sequence-specific bending and torsional potentials. Particularly, the sequence dependence is introduced by considering a simplified three-letters description, where the eventual presence
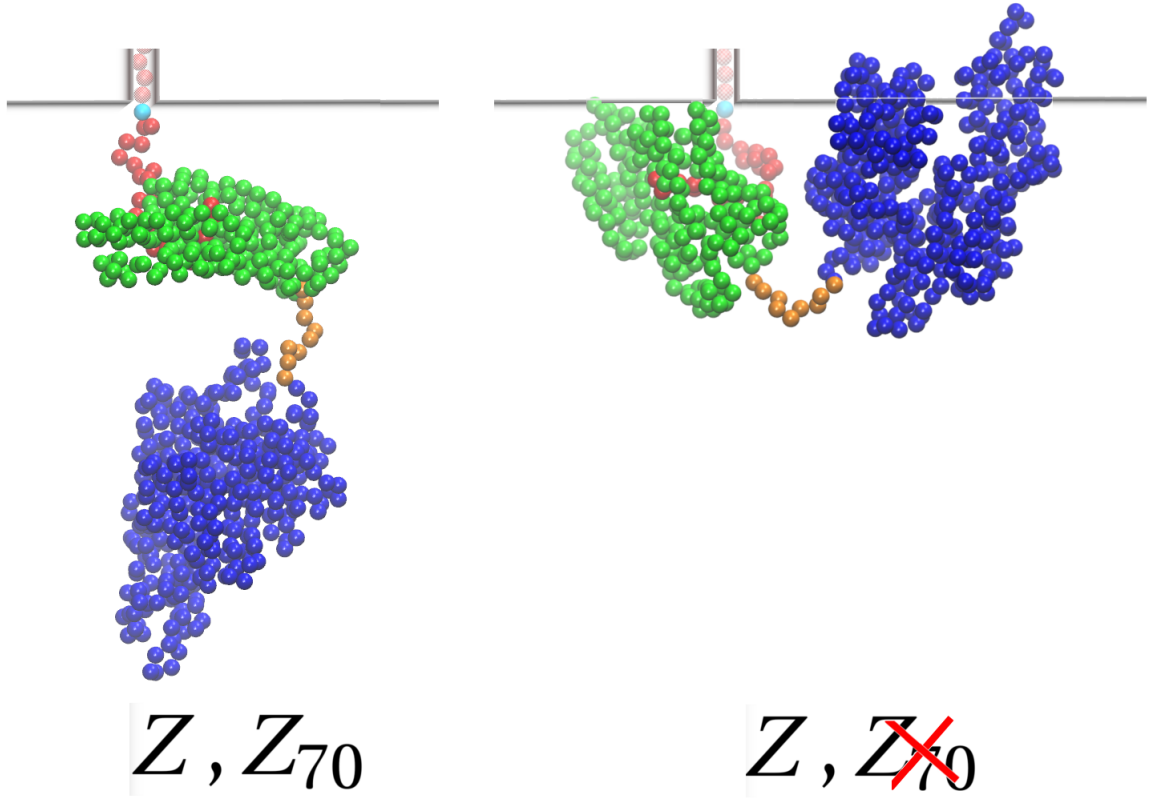
Figure 2.3 – Examples of coarse-grained structures, representing the same snapshots as in fig.2.2. While in the absence of Hsp70s both conformations contribute to $Z(n)$, upon chaperone binding the one on the right is not taken into account in $Z_{70}(n)$ due to sterical clash with the wall. The nth residue of the substrate, which in our MD simulations is constrained on the wall, is colored in cyan. The shaded beads inside the channel are here drawn only for representative purposes. Adapted from [112].

of glycine or proline amino acids is explicitly accounted for, while all the other residues are considered equivalent. In this context, for simplicity we focus on glycine- and proline-free substrates, thus making use of the functions denoted as O-X-Y and X-X in [113] for the bending and torsional contributions, respectively. The experimental structure of ADP-bound Hsp70 [93] (PDB: 2KHO) was used to model the chaperone. In particular, the NBD (residue 4 to 387) and SBD (residue 397 to 603) were treated as rigid bodies, while the flexibility of the interdomain linker was accounted for by means of the potential described above. In order to reproduce a correct chaperone-substrate arrangement, we took advantage of the substrate-bound X-ray structure of DnaK SBD (PDB: 1DKX [114]). MD simulations were performed using LAMMPS code [115] at constant temperature ($T = 300$ K) by means of a Langevin thermostat with damping parameter equal to 100 fs and using an integration timestep of 20 fs. For each value of n in the range $8 \leq n \leq 30$ we performed 16 independent MD simulations of $9 \cdot 10^8$ timesteps (examples of the convergence of the ratio $Z_{70}(n)/Z(n)$ are reported in fig.2.4), and
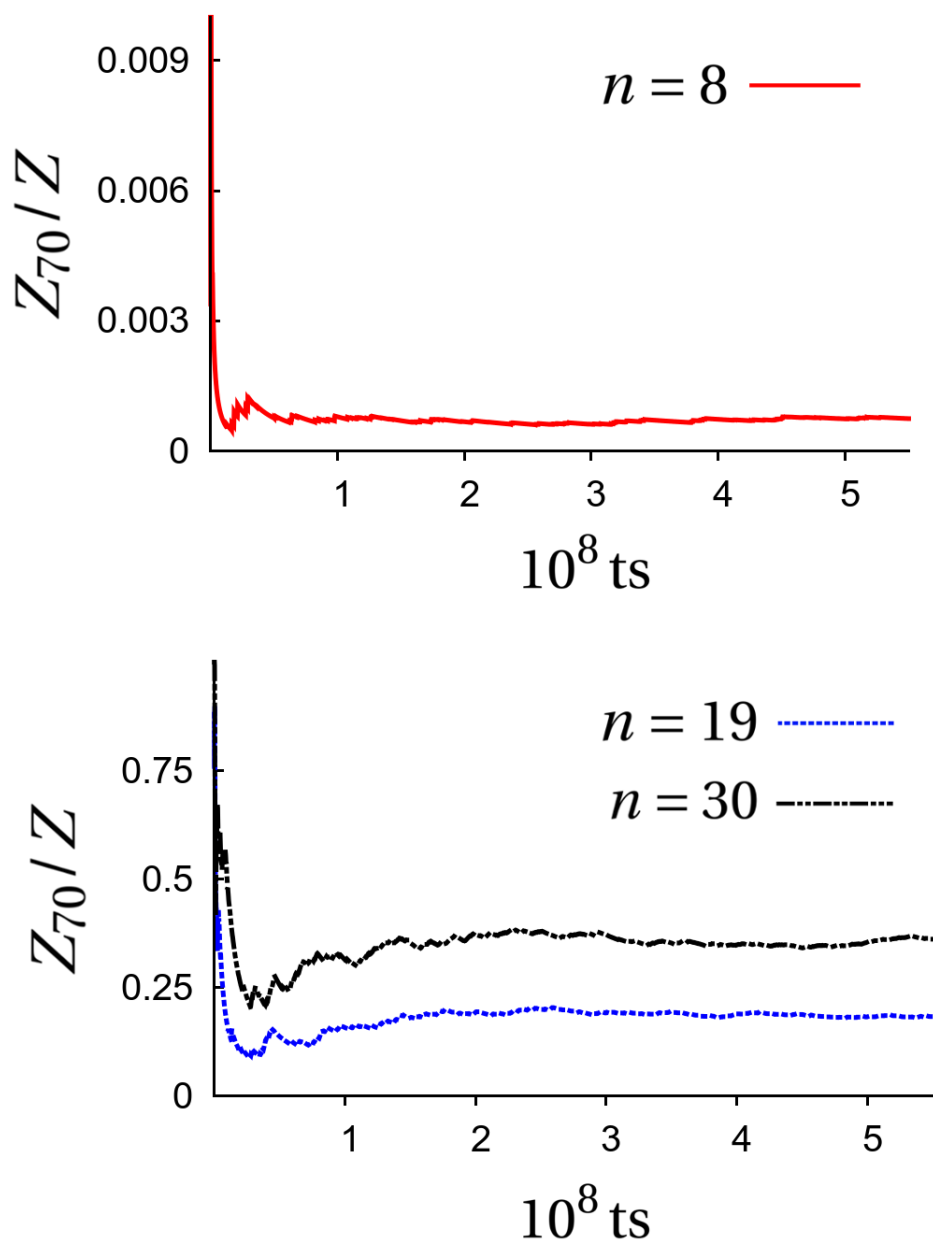
Figure 2.4 – Evolution of the ratio $Z_{70}(n)/Z(n)$ as a function of MD timesteps, for $n = 8$ (top panel, red continuous curve), $n = 19$ (bottom panel, blue dotted curve) and $n = 30$ (bottom panel, black dot-dashed curve).

the error on the free-energy profile was estimated by computing the standard deviation on the final values of the ratios.

From the computed values of $Z_{70}(n)/Z(n)$, we could retrieve $\Delta F_c(n)$ as a function of n by means of equation (2.6). The resulting free-energy landscape is reported in fig.2.5. As expected,
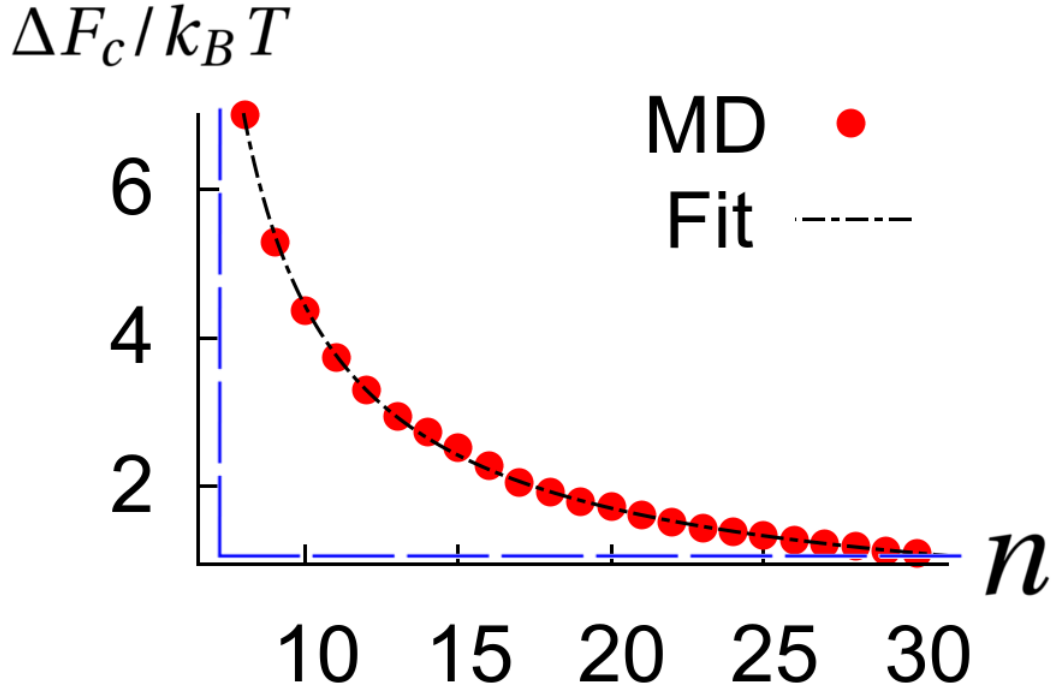
Figure 2.5 – Free-energy profile due to chaperone binding as a function of n (red circles). Error bars are smaller than the size of the symbols. The black dot-dashed line is obtained fitting the MD data with a power-law function $q(n-s)^{-t}$ by tuning the parameters q, s and t. The final values for the parameters are $q = 16.7 \pm 0.9, s = 5.2 \pm 0.1, t = 0.85 \pm 0.02$. The dashed blue line depicts the free-energy landscape predicted by the original Brownian ratchet, where the only effect of the chaperone is to prevent retrotranslocation beyond the binding site (infinite wall).

shorter imported fragments resulted into a larger fraction of rejected conformations, *i.e.* larger values of ΔF_c, thus leading to a free-energy gradient favoring the import of the protein. For comparison, we report the corresponding free-energy landscape predicted by the Brownian ratchet as it was originally proposed (blue dashed line). In that case, a flat profile was expected, with an infinite wall in correspondence of the binding site (this wall is located at $n = 8$ since we found no sterically-allowed conformations for lower values of n).

From the prediction based on more simplified assumptions, equation (2.7), it seems natural to wonder whether the free-energy landscape retrieved from MD simulations can be as well described by a power law. To this aim, we fitted the data by considering the function $q(n - s)^{-t}$ and adjusting the parameters q, s and t. The resulting function is plotted as a black dot-dashed line in fig.2.5 and describes very well the simulation data. The values of the parameters minimizing the squared displacement from the data are $q = 16.7 \pm 0.9, s = 5.2 \pm 0.1, t = 0.85 \pm 0.02$. Interestingly, these numbers are qualitatively in line with the "spherical cow" prediction given in equation (2.7). Indeed, the exponent t is not far from the predicted value 1, though definitely not in quantitative agreement with it. Moreover, by comparing the

value of q with the prefactor in equation (2.7), we can compute the equivalent chaperone radius, which results equal to about 40 Å. This value is twice the quantity used above, in agreement with the deliberate underestimation made in [107] and with the typical large values of the excluded volume of anisotropic objects when compared to spheres of similar size [27]. The qualitative agreement between the two different levels of description suggests that comparable thermodynamic forces could be obtained by the same entropic pulling mechanism for macromolecules of similar size.

2.2.2.1 Measuring the thermodynamic force

From the free-energy landscape reported in fig.2.5, we can give an estimate of the typical pulling force. Specifically, since the import of a bead increases by $\sigma = 3.8$ Å the contour length of the imported substrate, a simple approximation of the force exterted by the chaperone f_{ep} is given by considering the change in ΔF_c upon the import of a bead:

$$f_{\mathrm{ep}}\left(n + \frac{1}{2}\right) = -\frac{\Delta F_c(n+1) - \Delta F_c(n)}{\sigma} \, .$$

$$(2.8)$$

In the previous formula, the independent variable on the left-hand side is chosen to be equal
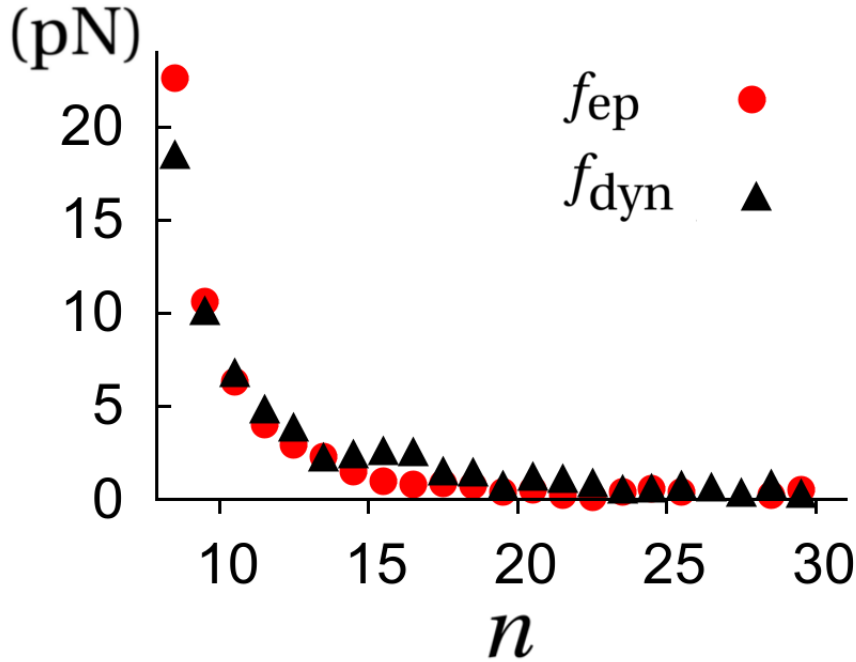

Figure 2.6 – Comparison between the force computed from the free-energy landscape f_{ep} (red circles), equation (2.8), and the postprocessed force f_{dyn} retrieved directly from the second set of MD simulations (black triangles). In both cases the error bars are about the same size as the symbols.

to the midpoint between n and $n+1$, since (looking at n as a continuous variable) the right-hand side is equal to the two-points approximation of the derivative of ΔF_c in $n+1/2$ [116]. The resulting force is reported in fig.2.6 (red circles). As shown in the plot, f_{ep} is in the piconewton range, starting from around 22 pN and decreasing as n increases.

MD simulations give also the possibility to more directly probe the thermodynamic effect of Hsp70 on the substrate, *e.g.* by measuring the average force acting on the fixed bead of the substrate (bead depicted in cyan in fig.2.3). To this aim, we measured the average force exerted on the fixed bead in a second set of MD simulations, where both the substrate and DnaK feel the presence of the wall. In other words, conformations such as the one shown in the right panel in fig.2.3 are not explored anymore by the system. The other simulation details were the same as in the previous section. Moreover, we performed analogous simulations where only the substrate was present (see below). In fig.2.7 we report some examples showing the convergence of the average value of the measured forces as a function of simulation time.
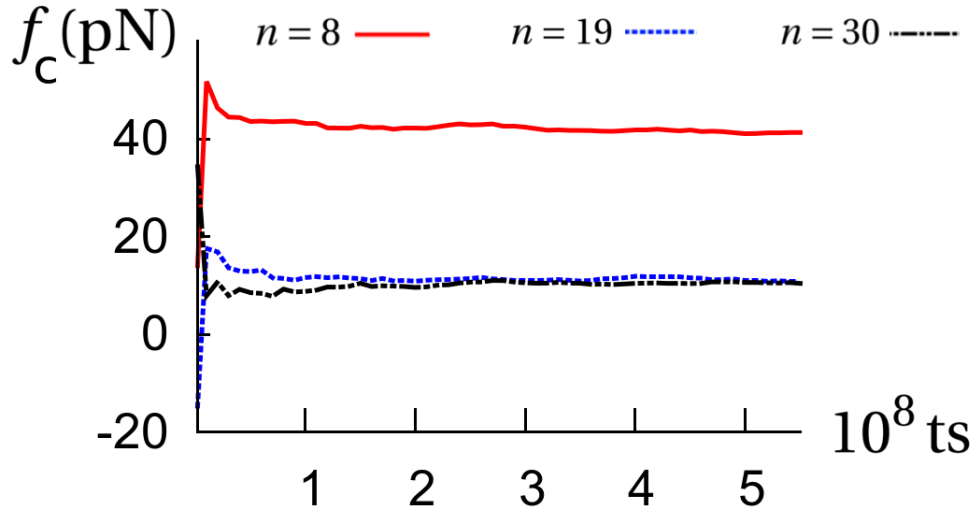
Figure 2.7 – Evolution of the force f_c acting on the fixed bead in the presence of the chaperone as a function of MD timesteps, for $n = 8$ (red continuous curve), $n = 19$ (blue dotted curve) and $n = 30$ (black dot-dashed curve).

How do the measured forces relate with f_{ep}? We note that the latter refers to the import process along the coordinate n. As for the former, a net average force is naturally expected to act on the fixed bead in the direction perpendicular to the wall due to the asymmetry of the system, both in presence (f_C) and absence (f_0) of chaperone. However, this force corresponds to the derivative of the free-energy (with fixed n) as a function of the coordinate x that measures the minimum distance between the fixed bead and the wall ($x = 0$ in all our MD simulations). More precisely, let us consider an extended ensemble where both the number of imported residues n and the minimum distance between the fixed bead and the wall x can be tuned. Let $\mathscr{F}_C(n, x)$ be the free energy corresponding to a given couple of values of n, x in the presence of

the chaperone, and let $\mathscr{F}_0(n, x)$ be its counterpart without DnaK. The average forces f_C and f_0 are thus given by

$$f_C(n) \equiv -\left.\frac{\partial \mathscr{F}_C(n, x)}{\partial x}\right|_{x=0}, \qquad f_0(n) \equiv -\left.\frac{\partial \mathscr{F}_0(n, x)}{\partial x}\right|_{x=0}. \tag{2.9}$$

Moreover, by construction $\Delta F_c(n) = \mathscr{F}_C(n, 0) - \mathscr{F}_0(n, 0)$. The force $f_C - f_0$ is thus not exactly equal to the import thermodynamic force f_{ep}, which corresponds to the variation in the chaperone free-energy contribution $\Delta F_c(n)$ upon the import of a residue. Nevertheless, we note that the conformational space explored by a substrate made of n residues and having the fixed bead at a minimum distance $x = \sigma$ from the wall is approximately the same as the conformational space of a substrate with $n+1$ residues, $x = 0$ and the last *two* beads fixed. It is reasonable to assume that the conformational free energy in this case is approximately equal to the one of the "normal" substrate of length $n+1$, *i.e.* when $x = 0$ and only the last bead is frozen. Within the extended ensemble, the previous statement can be written as

$$\mathscr{F}_C(n, \sigma) \simeq \mathscr{F}_C(n+1, 0), \tag{2.10}$$

with a similar formula holding for $\mathscr{F}_0$. Within this approximation, we can thus compute $\Delta F_c(n+1)$ starting from $\Delta F_c(n)$ and integrating the force difference $f_c - f_0$ along x:

$$\Delta F_c(n+1) = \mathscr{F}_C(n+1, 0) - \mathscr{F}_0(n+1, 0) \simeq \mathscr{F}_C(n, \sigma) - \mathscr{F}_0(n, \sigma) =$$

$$= \left[\mathscr{F}_C(n, 0) - \int_0^\sigma f_c(n, x)\, dx\right] - \left[\mathscr{F}_0(n, 0) - \int_0^\sigma f_0(n, x)\, dx\right] =$$

$$= \Delta F_c(n) - \int_0^\sigma \left[f_c(n, x) - f_0(n, x)\right] dx \tag{2.11}$$

Substituting the previous formula in equation (2.8), we thus conclude that f_{ep} is approximately equal to the path-averaged force $f_C - f_0$:

$$f_{\mathrm{ep}}\left(n + \frac{1}{2}\right) \simeq \frac{1}{\sigma} \int_0^\sigma \left[f_c(n, x) - f_0(n, x)\right] dx. \tag{2.12}$$

Nevertheless, in our simulations we have access only to the quantities $f_c(n, 0)$ and $f_0(n, 0)$. Therefore, we further approximate the right-hand side of the previous formula by considering the arithmetic mean between the forces in n and $n+1$. As a final result, we thus have

$$f_{\mathrm{ep}}\left(n + \frac{1}{2}\right) \simeq f_{\mathrm{dyn}}\left(n + \frac{1}{2}\right), \tag{2.13}$$

where we have defined

$$f_{\mathrm{dyn}}\left(n + \frac{1}{2}\right) \equiv \frac{f_c(n, 0) + f_c(n+1, 0)}{2} - \frac{f_0(n, 0) + f_0(n+1, 0)}{2}. \tag{2.14}$$

In fig.2.6 we check the goodness of the approximation provided by equation (2.13) by plotting f_{ep} (red circles) and f_{dyn} (black triangles) as a function of n. The comparison shows that the two quantities are in good agreement with each other, though for low values of n the value of f_{ep} results slightly larger than the corresponding f_{dyn}.

2.2.3   Stochastic simulations of the import process

The entropic-pulling free energy computed in the previous section enables us to address the problem of protein import into organelles as a diffusion of a particle on a n-dependent free-energy landscape. Protein import into organelles has been previously modeled as a one-dimensional stochastic process in the space of the imported residues [26, 117, 118]. In the present context, this protocol is justified by the timescale separation among substrate conformational dynamics, chaperone binding/unbinding and overall import. Indeed, the typical reconfiguration time of an unfolded protein (~ 100 ns [119]) is extremely fast compared to the experimentally-determined timescale for protein import into mitochondria (order of minutes [120]). Effects arising from substrate conformational dynamics, such as the chaperone-induced entropy reduction, can be thus conveniently represented as effective free-energy profiles influencing the import dynamics. Moreover, the import timescale is also significantly slower than chaperone binding but faster than chaperone unbinding at physiological conditions. Indeed, combining the Hsp70-peptide association rate of a chaperone in the ATP-bound state reported in table 2.1 (4.5×10^5 M^{-1}s^{-1} [97]) with the concentration of mtHsp70 found *in vivo* (70 μM [121]), we obtain a timescale for association equal to $\sim 10^{-2}$ s. Moreover, dissociation takes place from the ADP-bound state, over timescales $\sim 10^3$ s (see table 2.1 and references [78, 97, 122]). In the mitochondrial matrix there are nucleotide exchange factors that prompt the release of ADP and the ensuing rebinding of ATP, resulting in an acceleration of the dissociation of the substrate. Nonetheless, this likely applies to Hsp70s that are bound to the translocating polypeptide far from the pore, whose contribution to the free-energy profile is anyway negligible (compare fig.2.5). In the proximity of the pore, instead, the concomitant binding of the chaperones to TIM44 and the fast hydrolysis rate due to the presence of a pore-associated J-domain protein ensure fast, strong and, on average, long-lived chaperone binding. This suggests that, to our purposes, we can assume that a chaperone immediately and irreversibly binds each exposed binding site as soon as it is imported.

As a consequence, for the present purposes the number n_{in} of substrate residues that have been imported into the mitochondrial matrix is a convenient coordinate to describe the system, whose dynamics can be modeled as a diffusion process on the corresponding free-energy landscape[1]. The effective free-energy profile guiding the system evolution, F_{import}, results from protein unfolding [123] and active chaperone pulling [120]. As for the former, assuming a two-states folding behavior, a convenient choice to model the unfolding contribution to the

---

[1]The use of a different notation is due to an important difference with respect to the coordinate n used in the previous section. While n_{in} is the *total* number of imported residues, and is then used to monitor the translocation process, the coordinate n is related to a specific chaperone and gives the contour distance between the binding point and the membrane pore.

free-energy landscape is a tunable sigmoidal function $F_u(n_{\text{in}})$

$$F_u(n_{\text{in}}) = \frac{F_u^{\max}}{1 + \exp\left[5 - 10(n_{\text{in}} - 10)/\delta n\right]}, \tag{2.15}$$

where n_{in} is the total number of imported residues, and $F_u^{\max}$ and δn are tunable parameters. In fig.2.8 we report the plots of some representative examples of equation (2.15). As can be seen from the graph, the two parameters measure the total unfolding free energy ($F_u^{\max}$) and the cooperativity of the unfolding process (δn), with smaller values of δn corresponding to higher cooperativity. By tuning these parameters, the formula can account for the wide variety of imported proteins [124]. The pulling action of the chaperone was modeled taking
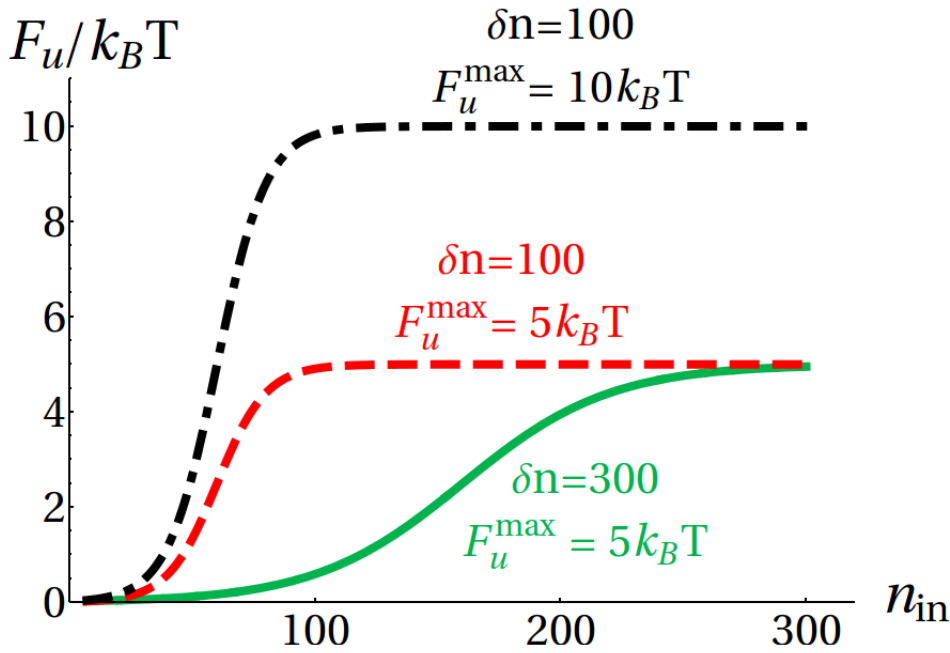
Figure 2.8 – Influence of the parameters $F_u^{\max}$ and δn on the unfolding free energy. Reprinted from [112].

advantage of the free-energy profile determined from the molecular simulations reported in the previous section. Particularly, we assumed here that: i) Hsp70s associate with each binding site as soon as it emerges from the pore, since they are targeted at the TIM pore exit by specific interactions [88]; ii) we considered only the contribution arising from the Hsp70 closest to the pore, taking into account the relatively fast decrease of the slope of ΔF_c (see fig.2.5) and the average frequency of binding sites (one every 35 amino acids [96]). Therefore, we added to the unfolding free-energy $F_u(n_{\text{in}})$ the chaperone contribution $\Delta F_c(n_{\text{in}} - n_B)$, with n_B corresponding to the position of the binding site closest to the pore, measured from the matrix terminus of the substrate.

The import process was simulated by means of a Monte Carlo (MC) algorithm driven by the free-energy landscape F_{import}. For a system at position n_{in}, a trial move was attempted to either $n_{\text{in}} + 1$ or $n_{\text{in}} - 1$ with equal probability and accepted according to the Metropolis criterion based on the free energy F_{import}. To capture the sequence heterogeneity of the proteome, for each choice of F_u^{max} and δn we generated 25 independent binding-site distributions, with the sole prescription that the average distance between consecutive binding sites was 35 residues as indicated by experiments [96]. For every distribution we performed 10 independent realizations of the import process. Average import times were estimated by counting the total number of MC timesteps needed for the translocation process to be completed. This protocol is justified by the fact that MC simulations correspond to overdamped Langevin dynamics when only local moves are considered [125, 126]. Rescaling the obtained import times by the acceptance rate, as proposed in [127], did not affect the results, because of the large fraction of accepted moves observed in all the simulations ($> 95\%$).
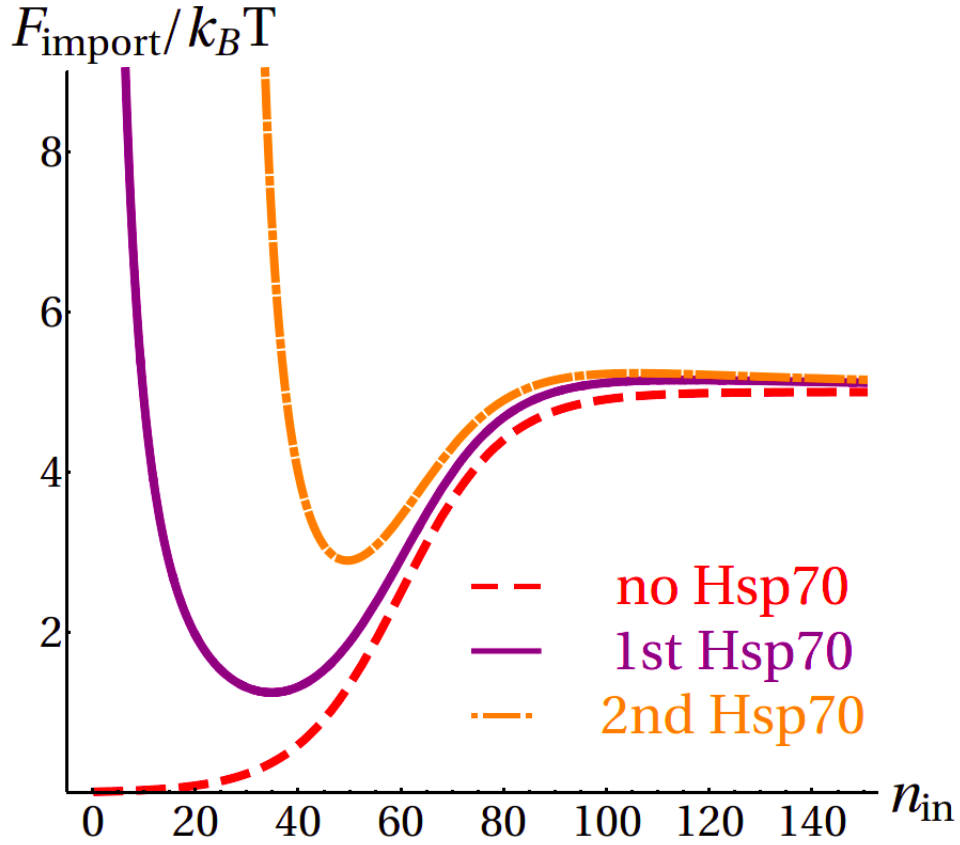
Figure 2.9 – Evolution of the total free-energy F_{import} in a representative import process. Reprinted from [112].

As an example, in fig.2.9 we illustrate the evolution of the free-energy landscape during the import process of a protein with $F_u^{\text{max}} = 5 k_B T$, $\delta n = 100$ and two binding sites at $n_B = 0$ (*i.e.* at the matrix terminus) and $n_B = 28$. At the beginning of the import process, no chaperone is

bound to the substrate and the import free energy is simply given by $F_{\text{import}}(n_{\text{in}}) = F_u(n_{\text{in}})$ (red dashed curve). As soon as the first binding site is imported, a chaperone molecule binds the substrate and its contribution ΔF_c is added to $F_u(n_{\text{in}})$ starting from the binding site $n_B = 0$: $F_{\text{import}}(n_{\text{in}}) = F_u(n_{\text{in}}) + \Delta F_c(n_{\text{in}})$ (purple continuous curve). Finally, after the second binding site ($n_B = 28$) is imported, another chaperone binds and the resulting free energy is $F_{\text{import}}(n_{\text{in}}) = F_u(n_{\text{in}}) + \Delta F_c(n_{\text{in}} - 28)$ (orange dot-dashed curve).
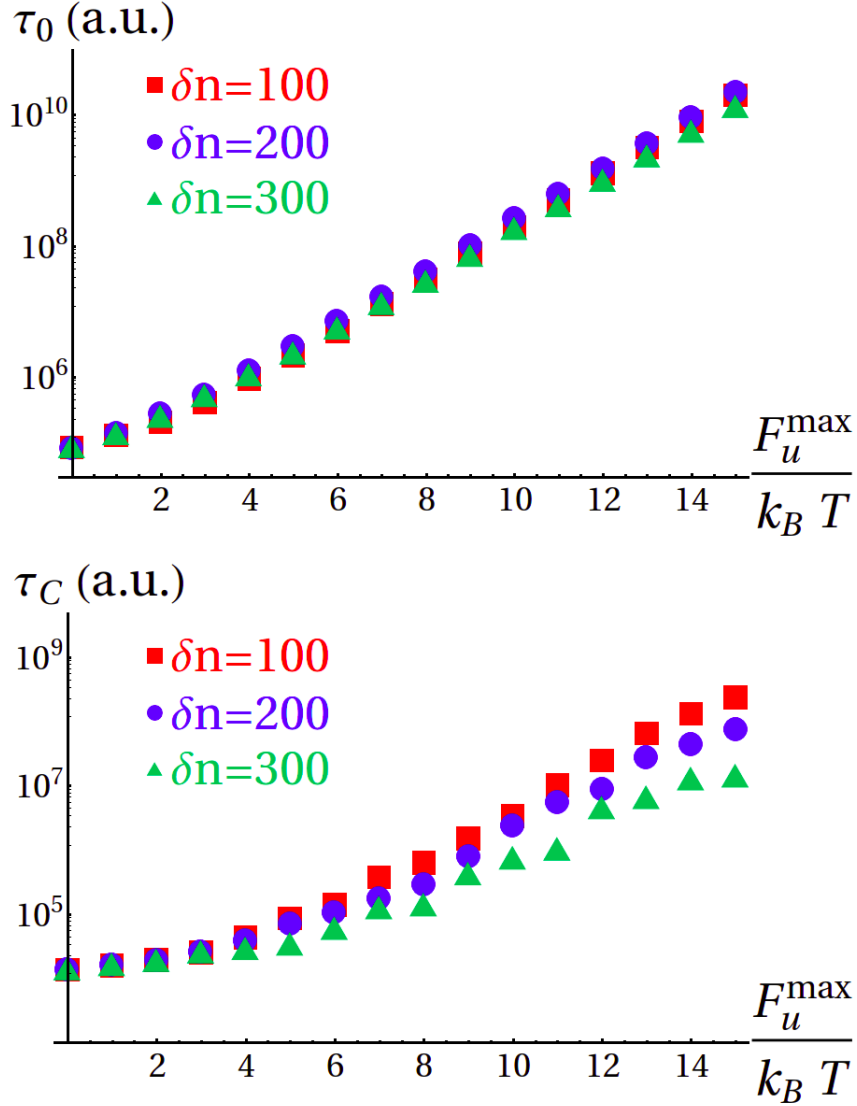
Figure 2.10 – Top: average import times in the absence of chaperone (τ_0) as a function of F_u^{max} for different cooperativities (values for $F_u^{\text{max}} \geq 12 k_B T$ were extrapolated by fitting the data in the range $4 k_B T \leq F_u^{\text{max}} \leq 11 k_B T$ with exponential functions). Bottom: average import times in the presence of Hsp70 (τ_C) for the same cases as in the top panel. Reprinted from [112].

Following this approach, we computed the average import time of 300-residue proteins [111]

for different values of δn and a range of F_u^{max} corresponding to the stability of a large fraction of the proteome [128]. In absence of Hsp70 assistance, the system must invariably overcome a free-energy barrier, and the average import time τ_0 increases exponentially with F_u^{max}, independently of cooperativity (fig.2.10 top). In all the considered cases, the average import time for the chaperone-assisted process, τ_C, is sensibly smaller than τ_0 (fig.2.10 bottom). The chaperone pulling force reduces but does not completely eliminate the unfolding free-energy difference for stable proteins (large F_u^{max}), as in the case of the representative process shown in fig.2.9. In this regime, the import is still an activated process, and the average times increase exponentially with F_u^{max}. Conversely, the pulling action of Hsp70 dominates over the unfolding contribution for marginally stable proteins (small F_u^{max}), thus resulting in values of τ_C comparable to what found for the extreme case $F_u^{\text{max}} = 0$. The import kinetics is further modulated by δn, with high cooperativity (small δn) resulting in longer translocation times.
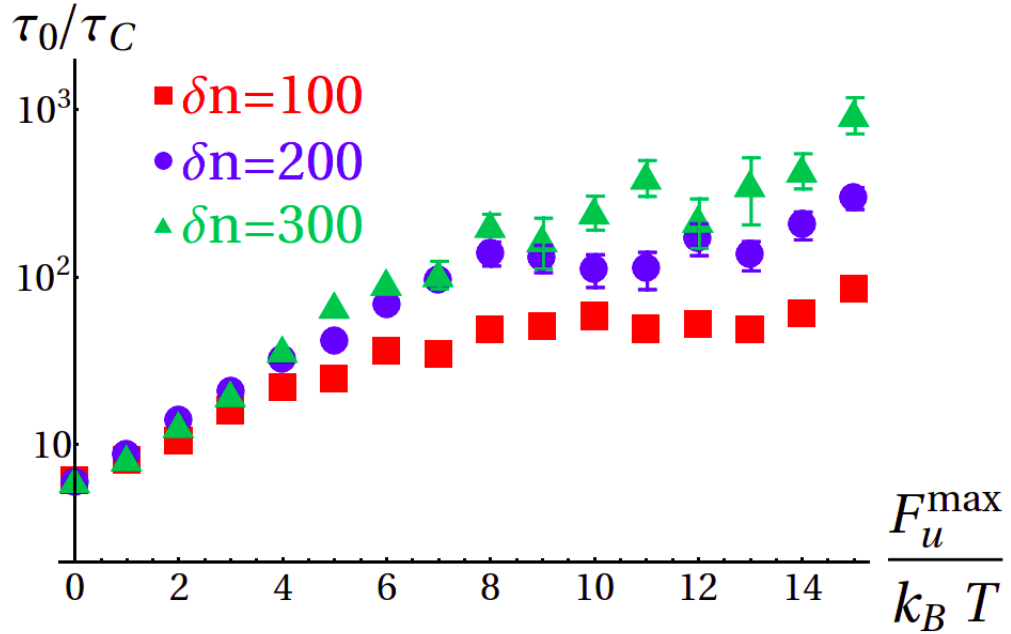
Figure 2.11 – Acceleration of the process due to the assistance of Hsp70, expressed as the ratio τ_0/τ_C for the same cases as in fig.2.10. Reprinted from [112].

In fig.2.11 we illustrate the chaperone-induced kinetic advantage by reporting the ratio τ_0/τ_C. This ratio ranges from a 10-fold gain for marginally stable proteins to 10^3 for extremely stable and noncooperative substrates, with the majority of the proteome ($F_u^{\text{max}} \geq 8k_B T$ [128]) accelerated at least 100 times. If we take into account that protein import into mithocondria has been measured to happen in the timescale of several minutes [120], our model indicates that the translocation process in the absence of chaperones would probably extend to hours or days. Since such a slow process would clearly be incompatible with the average lifespan of proteins and the duration of the cell cycle, our results provide a molecular basis to support the essential role of chaperones in the *in vivo* import process.

2.3 Conclusions

To summarize, in this chapter we derived a free-energy profile for the import process based on a molecular description of Hsp70 that rationalizes the requirement for chaperone assistance in mitochondrial protein import observed in experiments. The present results can be applied to other cases of Hsp70-driven translocation, namely protein import into ER [89] and chloroplasts [90]. In the ER case, in particular, the pore is much simpler than in mitochondria, as it spans just a single membrane. The presence of Sec63, a pore-associated protein containing a J domain, ensures also in this context that the extended ATP-driven Hsp70 ultra-affinity prevails in the competition against other translocation counterproductive interactions [129]. Moreover, this approach based on the combination of molecular simulations and kinetic modeling can be easily extended to other Hsp70-mediated cell processes. In particular, this free-energy picture could help to understand some recent results pointing towards a fundamental role of Hsp70 in preventing the stalling of translation at ribosomes [91, 92]. Owing to the universality of the interaction responsible for the effects studied here, namely excluded volume, the same principles could apply to similar processes driven by other biomolecules.

3 The shape of a stretched polymer

3.1 Polymers in a good solvent have anisotropic shapes

3.1.1 Polymers adapt their shape to the external conditions

Polymer chains are by definition strongly-fluctuating objects. Due to their soft nature, they do not have well-defined shapes, but rather adapt their conformational ensembles to the environment surrounding them. At the same time, the shape of a polymer affects the way it interacts with other molecules in solution, *e.g.* quantitatively determining the excluded volume of the chain, thus ultimately influencing the thermodynamic properties of the solution [27]. In a poor solvent, where monomer-monomer and monomer-solvent interactions result into a net attraction between different parts of the chain [27], polymers collapse into a roughly spherical shape, as seen for example in the case of single-chain globular proteins [130].

In the case of θ- or good-solvent conditions, where the free energy is dominated by conformational entropy [27], conformations can fluctuate more wildly and the shape of a polymer is more sensitive to the environment. In the last two decades, many theoretical and experimental works have investigated how the shape depends on several factors such as confinement [131–133], topology [134–136] or crowding [130, 137–139], which are relevant to many cellular processes.

In vivo biopolymers are often subjected to mechanical forces. For example, during replication DNA is repeatedly pulled and twisted by enzymes [140]. Although it is expected that a highly-stretched chain behaves as a stiff rod, a proper account for the effect of an external pulling force on the shape of a polymer has been lacking, even at the simplest level. Here we compute, by means of both analytical calculations and Monte Carlo (MC) simulations, some key global quantities characterizing the shape of a stretched chain. Before deriving and discussing our findings, we dedicate the next two sections to some basic results on unperturbed FJCs. Apart from being instructive about the physics underlying polymers anisotropy, they will also serve to introduce the toolbox which will be used to characterize the shape properties of stretched chains.

3.1.2 Random walks and symmetry breaking

Strikingly, when it comes to shape properties even the simplest models for polymers give non-trivial results. For example, let us consider the Freely Jointed Chain (FJC) model[1], a useful representation of a polymer in a good solvent where the conformation of a chain made by $N+1$ monomers is described as a three-dimensional random walk $\{\boldsymbol{R}_i\}_{i=0}^{N}$, where the segments $\boldsymbol{r}_i \equiv \boldsymbol{R}_i - \boldsymbol{R}_{i-1}$ have all equal length l [27]. As already noticed in an early work by Kuhn [141], in spite of the spherical symmetry of the system, the typical shape of an unperturbed FJC is an elongated ellipsoid, whose random orientation preserves the overall symmetry. This finding can result a little awkward at first sight, but can be easily rationalized analyzing the basic features of the FJC. In this regard, let us focus on the end-to-end vector $\boldsymbol{R}_e$, which is defined as

$$\boldsymbol{R}_e \equiv \boldsymbol{R}_N - \boldsymbol{R}_0 = \sum_{i=1}^{N} \boldsymbol{r}_i \,. \tag{3.1}$$

The FJC can be mapped on a random walk where the chain length N plays the role of time [16]. The end-to-end vector is therefore distributed according to a Gaussian formula

$$P_N(\boldsymbol{R}_e) = \left(\frac{3}{2\pi N l^2}\right)^{\frac{3}{2}} e^{-\frac{3\boldsymbol{R}_e^2}{2Nl^2}} \tag{3.2}$$

Because of the spherical symmetry of the system, one has $\langle \boldsymbol{R}_e \rangle = 0$, where $\langle \cdots \rangle$ denotes averaging over the conformational ensemble [27]. However, its squared average gives

$$\langle \boldsymbol{R}_e^2 \rangle = Nb^2 \,. \tag{3.3}$$

Moreover, this value also corresponds to the most probable length of $\boldsymbol{R}_e$. Therefore, it is only due to an average over all the orientations that one obtains a null end-to-end vector. Thus, one may already suspect that for a given orientation of $\boldsymbol{R}_e$ the spatial distribution of monomers is on average more elongated in the direction of the end-to-end vector.

To get a better insight, let us focus on a typical realization, *i.e.* let us restrict the conformational ensemble to the walks satisfying with $\boldsymbol{R}_e^2 = Nb^2$. Since we are interested in the shape properties of the FJC, we select an arbitrary orientation of $\boldsymbol{R}_e$ and choose a reference frame with the origin on the first monomer and whose z axis is oriented along the end-to-end vector. For a walk with a fixed $\boldsymbol{R}_e$, it can be shown that the probability distribution $P_N^k(\boldsymbol{R}_k)$ for the position vector of monomer k is equal to [27]

$$P_N^k(\boldsymbol{R}_k) = \left(\frac{3}{2\pi\sigma_k^2}\right)^{\frac{3}{2}} e^{-\frac{3\left(\boldsymbol{R}_k - \frac{k}{N}\boldsymbol{R}_e\right)^2}{2\sigma_k^2}} \,, \tag{3.4}$$

---

[1]We have already introduced the FJC in section 2.2.1, but we repeat here its definition for convenience.

where

$$\sigma_k^2 \equiv \frac{k(N-k)}{N} l^2 \,. \tag{3.5}$$

In other words, the partial end-to-end vector is still distributed according to a Gaussian formula, but with renormalized average and variance. Note that substituting $k = 0, N$ we retrieve the fixed ends of the polymer with null variance. We now compute the mean squared radius of gyration $\langle R_g^2 \rangle$ of the constrained FJC, which is a widely-used estimator of the size of a polymer [27]. Assuming all monomers to have equal mass, by definition we have

$$R_g^2 \equiv \frac{1}{N+1} \sum_{i=0}^{N} (\boldsymbol{R}_i - \boldsymbol{R}_{cm})^2 \,, \tag{3.6}$$

where $\boldsymbol{R}_{cm} \equiv \sum_i \boldsymbol{R}_i /(N+1)$ is the position of the center of mass of the chain. By rearranging the sum in the right-hand side of equation (3.6), it can be easily shown that the Lagrange formula holds [27]:

$$R_g^2 = \frac{1}{2(N+1)^2} \sum_{i=0}^{N} \sum_{j=0}^{N} \left(\boldsymbol{R}_i - \boldsymbol{R}_j \right)^2 \,. \tag{3.7}$$

For long chains, the sums in the previous formula can be approximated as integrals. As a result, the average of R_g^2 can be written as

$$\langle R_g^2 \rangle \simeq \frac{1}{2N} \int_0^N di \int_0^N dj \, \langle \left(\boldsymbol{R}_i - \boldsymbol{R}_j \right)^2 \rangle = \frac{1}{N} \int_0^N di \int_0^i dj \, \langle \left(\boldsymbol{R}_i - \boldsymbol{R}_j \right)^2 \rangle \,. \tag{3.8}$$

We recognize in the integral the squared average of the end-to-end vector of the subchain connecting monomers j and $i > j$. In an unconstrained chain, thanks to the statistical independence of different segments, one could directly substitute for the formula of a chain of length $i - j$ [27]. However, in this case the calculation is more involved, since because of the constraint nonoverlapping parts of the chain are coupled to each other and one has to use suitably-adapted versions of equation (3.4) according to the values of i and j. The detailed derivation is beyond our present purposes, and can be found in [142]. The final result is

$$\langle R_g^2 \rangle = \frac{1}{12} \left(1 + \frac{\boldsymbol{R}_e^2}{Nl^2} \right) \,. \tag{3.9}$$

Note that, by averaging over all the possible choices of the end-to-end vector, we retrieve the known result for linear chains

$$\langle R_g^2 \rangle = \frac{1}{6} Nl^2 \,, \tag{3.10}$$

while imposing $\boldsymbol{R}_e = 0$ gives the radius of gyration of a ring polymer $\langle R_g^2 \rangle = Nl^2/12$ [27].

To our aim, it is more instructive to consider the contributions to $\langle R_g^2 \rangle$ coming from the single

directions. Particularly, defining

$$\langle z_g^2 \rangle \equiv \frac{1}{N} \int_0^N di \int_0^i dj \, \langle (z_i - z_j)^2 \rangle \tag{3.11}$$

and considering similar formulas for the x and y directions, by construction we have

$$\langle R_g^2 \rangle = \langle x_g^2 \rangle + \langle y_g^2 \rangle + \langle z_g^2 \rangle \, . \tag{3.12}$$

The results from the single contributions are [142]

$$\langle x_g^2 \rangle = \langle y_g^2 \rangle = \frac{Nl^2}{36} \tag{3.13}$$

and

$$\langle z_g^2 \rangle = \frac{Nl^2}{36} \left(1 + \frac{3R_e^2}{Nl^2} \right) . \tag{3.14}$$

Comparing equation (3.13) and equation (3.14) shows that on average the radius of gyration receives the largest contribution from the direction along the end-to-end vector, thus suggesting a breaking of the spherical symmetry. Particularly, in the "typical" case where $R_e^2 = Nl^2$, $\langle z_g^2 \rangle$ is four times $\langle x_g^2 \rangle$, thus pointing towards a rather elongated shape for the polymer.

Nevertheless, the present analysis catches only approximately the actual extent of anisotropies of the chain and thus serves only as an intuitive argument explaining their origin. A proper quantitative description of these features requires more refined methods, which are introduced in the next section.

3.1.3   Inertia tensor and asphericity

The radius of gyration is a useful quantity to measure the spatial extent of a macromolecule. Nevertheless, it cannot capture the presence of anisotropies, which ultimately determine the overall shape of the polymer. In this respect, a more suitable tool is its natural extension provided by the inertia tensor $\mathscr{T}$, whose elements are by definition [143]

$$\mathscr{T}_{\alpha\beta} \equiv \frac{1}{N+1} \sum_{i=0}^N (\alpha_i - \alpha_{cm})(\beta_i - \beta_{cm}), \qquad \alpha, \beta \in \{x, y, z\} \, . \tag{3.15}$$

The analogue of equation (3.7) reads in this case

$$\mathscr{T}_{\alpha\beta} = \frac{1}{2(N+1)^2} \sum_{i=0}^N \sum_{j=0}^N (\alpha_i - \alpha_j)(\beta_i - \beta_j) = \frac{1}{(N+1)^2} \sum_{i=0}^N \sum_{j=0}^i (\alpha_i - \alpha_j)(\beta_i - \beta_j). \tag{3.16}$$

The eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ of $\mathscr{T}$ are proportional to the square of the semiaxes of the ellipsoid which best approximates the shape of the polymer, while the corresponding

eigenvectors describe its orientation. Moreover, by definition

$$R_g^2 = \lambda_1 + \lambda_2 + \lambda_3 \,. \tag{3.17}$$

Numerical computations have shown that the ensemble averages of the eigenvalues of an unperturbed FJC are in the ratios $\langle \lambda_1 \rangle : \langle \lambda_2 \rangle : \langle \lambda_3 \rangle = 11.8 : 2.7 : 1$ [143]. According to these results, the actual spatial distribution of the monomers is even more asymmetric than what expected from the analysis performed in the previous section: not only the symmetry is broken in all the three directions (*i.e.* also the cylindrical symmetry suggested by equation (3.13) is broken), but also the elongation is more pronounced than expected.

A useful index to quantify this anisotropy is provided by the asphericity [143, 144]

$$\mathscr{A} \equiv \frac{\sum_s \langle (\lambda_s - \bar{\lambda})^2 \rangle}{6 \langle \bar{\lambda}^2 \rangle} = \frac{3 \langle \mathrm{Tr} \left(\mathscr{T}^2 \right) \rangle}{2 \langle (\mathrm{Tr} \mathscr{T})^2 \rangle} - \frac{1}{2} \,, \tag{3.18}$$

where $\bar{\lambda} = \sum_s \lambda_s / 3$ is the arithmetic mean of the eigenvalues. For a perfectly symmetric distribution, all the eigenvalues have equal magnitude $\lambda_s = \bar{\lambda}$, so that $\mathscr{A} = 0$. In the opposite limit of a rod-like chain, one eigenvalue dominates over the others ($\lambda_1 \gg \lambda_2, \lambda_3$) and as a result $\mathscr{A} = 1$. The asphericity of an unperturbed FJC can be computed exactly, and it has been shown that $\mathscr{A} = 10/19 \simeq 0.53$, independently of N [143]. The original derivation in [143] implied the construction of a suitable generating function, which let the authors find some elegant and general results for an unrestricted random walk in $\mathscr{D}$ dimensions. However, for our present purposes it suffices to approach the problem in a simpler way. According to equation (3.18), in order to compute $\mathscr{A}$ we need to calculate the averages $\langle \mathrm{Tr} \left(\mathscr{T}^2 \right) \rangle$ and $\langle (\mathrm{Tr} \mathscr{T})^2 \rangle$, which can be written explicitly as a combination of quadratic terms of $\mathscr{T}$. Such terms can always be decomposed as sums of end-to-end distances of independent subportions of the chain, whose average can be computed by means of equation (3.3).

More in detail, let us start by writing explicitly $\langle \mathrm{Tr} \left(\mathscr{T}^2 \right) \rangle$ as

$$\langle \mathrm{Tr} \left(\mathscr{T}^2 \right) \rangle = \langle \mathscr{T}_{xx}^2 \rangle + \langle \mathscr{T}_{yy}^2 \rangle + \langle \mathscr{T}_{zz}^2 \rangle + 2 \langle \mathscr{T}_{xy}^2 \rangle + 2 \langle \mathscr{T}_{xz}^2 \rangle + 2 \langle \mathscr{T}_{yz}^2 \rangle \,. \tag{3.19}$$

The spherical symmetry of the system implies that $\langle \mathscr{T}_{xx}^2 \rangle = \langle \mathscr{T}_{yy}^2 \rangle = \langle \mathscr{T}_{zz}^2 \rangle$ and $\langle \mathscr{T}_{xy}^2 \rangle = \langle \mathscr{T}_{xz}^2 \rangle = \langle \mathscr{T}_{yz}^2 \rangle$, so that only two independent terms are present in the previous formula, which can be rewritten as

$$\langle \mathrm{Tr} \left(\mathscr{T}^2 \right) \rangle = 3 \langle \mathscr{T}_{xx}^2 \rangle + 6 \langle \mathscr{T}_{xy}^2 \rangle \,. \tag{3.20}$$

In order to compute explicitly the right-hand side of equation (3.20), we proceed as in the previous section. Particularly, for large N we can consider the continuum limit of equation (3.16) and write

$$\mathscr{T}_{\alpha\beta} \simeq \frac{1}{N^2} \int_0^N di \int_0^i dj \left(\alpha_i - \alpha_j \right) \left(\beta_i - \beta_j \right), \qquad \alpha, \beta \in \{x, y, z\} \,. \tag{3.21}$$

We thus have

$$\langle \mathcal{T}_{xy}^2 \rangle = \frac{1}{N^4} \int_0^N di \int_0^i dj \int_0^N di' \int_0^{i'} dj' \, \langle (x_i - x_j)(y_i - y_j)(x_{i'} - x_{j'})(y_{i'} - y_{j'}) \rangle \qquad (3.22)$$

and

$$\langle \mathcal{T}_{xx}^2 \rangle = \frac{1}{N^4} \int_0^N di \int_0^i dj \int_0^N di' \int_0^{i'} dj' \, \langle (x_i - x_j)^2 (x_{i'} - x_{j'})^2 \rangle . \qquad (3.23)$$

The averages in the previous equations can be computed by properly taking into account the overlap between the subchains of the polymer lying between j, i and j', i'. As a consequence, several cases have to be considered (obviously performing the swap $j \leftrightarrow j'$ and $i \leftrightarrow i'$ results into equivalent cases). Before delving into their explicit treatment, it is convenient to introduce the quantities

$$c_{ij} \equiv x_i - x_j, \qquad f_{ij} \equiv y_i - y_j . \qquad (3.24)$$

Recognizing $x_i - x_j$ and $y_i - y_j$ as the x- and y-component of the end-to-end vector of the subchain connecting monomers j and i, starting from equation (3.2) it is straightforward to show that

$$\langle c_{ij} \rangle = \langle f_{ij} \rangle = 0,$$

$$\langle c_{ij}^2 \rangle = \langle f_{ij}^2 \rangle = \frac{(j-i)l^2}{3},$$

$$\langle c_{ij}^3 \rangle = \langle f_{ij}^3 \rangle = 0,$$

$$\langle c_{ij}^4 \rangle = \langle f_{ij}^4 \rangle = \frac{(j-i)^2 l^4}{3}.$$

Moreover, we note that c_{ij} has the following property

$$c_{ij} = c_{ik} + c_{kj},$$

as a direct substitution in equation (3.24) shows (an analogous formula holds for f_{ij}).

Let us consider the integrand in equation (3.22). Because of the mutual independence of different projections of a random walk, we can write

$$\langle (x_i - x_j)(y_i - y_j)(x_{i'} - x_{j'})(y_{i'} - y_{j'}) \rangle = \langle (x_i - x_j)(x_{i'} - x_{j'}) \rangle \langle (y_i - y_j)(y_{i'} - y_{j'}) \rangle ,$$

which thanks to the spherical symmetry of the system can be rewritten as

$$\langle (x_i - x_j)(x_{i'} - x_{j'}) \rangle^2 . \qquad (3.25)$$

In order to compute the previous average, we have to consider three different cases:

- $j < i < j' < i'$

  In this case no overlap is present, thus

  $$\langle (x_i - x_j)(x_{i'} - x_{j'}) \rangle \equiv \langle c_{ij} c_{i'j'} \rangle = \langle c_{ij} \rangle \langle c_{i'j'} \rangle = 0 \tag{3.26}$$

- $j < j' < i < i'$

  In this case the two subchains overlap in the region between monomers j' and i. However, we note that we can rearrange the average as

  $$\langle c_{ij} c_{i'j'} \rangle = \langle (c_{ij'} + c_{j'j})(c_{i'i} + c_{ij'}) \rangle = \langle c_{ij'} c_{i'i} + c_{j'j} c_{i'i} + c_{ij'}^2 + c_{j'j} c_{ij'} \rangle = \frac{i - j'}{3} l^2, \tag{3.27}$$

  where in the third step we took advantage of the statistical independence of all the factors involved in each term.

- $j < j' < i' < i$

  In this case the two subchains overlap in the region between monomers j' and i'. Proceeding as above, we can write

  $$\langle c_{ij} c_{i'j'} \rangle = \langle (c_{ii'} + c_{i'j'} + c_{j'j}) c_{i'j'} \rangle = \langle c_{ii'} c_{i'j'} + c_{i'j'}^2 + c_{j'j} c_{i'j'} \rangle = \frac{i' - j'}{3} l^2. \tag{3.28}$$

Plugging equations (3.26), (3.27) and (3.28) into equation (3.22) finally gives

$$\langle \mathcal{T}_{xy}^2 \rangle = \frac{1}{810} N^2 l^4. \tag{3.29}$$

A similar computation can be performed to calculate $\langle \mathcal{T}_{xx}^2 \rangle$, whose final result is

$$\langle \mathcal{T}_{xx}^2 \rangle = \frac{1}{180} N^2 l^4. \tag{3.30}$$

Substituting equations (3.30) and (3.29) into equation (3.20), we find

$$\langle \text{Tr}\left(\mathcal{T}^2\right) \rangle = \frac{23}{540} N^2 l^4. \tag{3.31}$$

By following the same procedure, we can also compute $\langle (\text{Tr}\,\mathcal{T})^2 \rangle$. The final result is

$$\langle (\text{Tr}\,\mathcal{T})^2 \rangle = \frac{19}{540} N^2 l^4. \tag{3.32}$$

Plugging equations (3.31) and (3.32) into equation (3.18), we finally find that the asphericity of an unperturbed FJC is independent of N and is equal to

$$\mathscr{A} = \frac{10}{19} \simeq 0.53, \tag{3.33}$$

in agreement with the result reported in [143].

3.2 The shape of a polymer under tension

3.2.1 Basic properties of a stretched chain

In the present section, we make use of the tools introduced above to characterize the shape of a stretched FJC. We first address the exact computation of the probability distribution of its end-to-end vector [145], since many subsequent results are derived from it. At equilibrium, the conformations of the chain are distributed according to the stretching energy $-\boldsymbol{f} \cdot \boldsymbol{R_e} = -\boldsymbol{f} \cdot \sum_i \boldsymbol{r_i}$, where $\boldsymbol{f}$ is the applied external force. Since the force is coupled independently to each segment, the joint probability distribution of $\{\boldsymbol{r_i}\}$ can be factorized as $\prod_i p(\boldsymbol{r_i})$, where

$$p(\boldsymbol{r_i}) = \frac{e^{\beta \boldsymbol{f} \cdot \boldsymbol{r_i}} \delta(|\boldsymbol{r_i}| - l)}{q}. \tag{3.34}$$

In the previous formula $\beta \equiv 1/k_B T$, where k_B is the Boltmann's constant and T the temperature, and

$$q = \int d\boldsymbol{r}\, e^{\beta \boldsymbol{f} \cdot \boldsymbol{r_i}} \delta(|\boldsymbol{r_i}| - l) = 4\pi l^2 \frac{\sinh(\beta f l)}{\beta f l} \tag{3.35}$$

is the partition function of a single segment. Starting from q and choosing a reference frame with the z axis oriented along the force, a straightforward computation leads to the well-known result [27]

$$\langle \boldsymbol{r} \rangle = l \mathcal{L}(v) \hat{\boldsymbol{e}}_z, \tag{3.36}$$

where $\hat{\boldsymbol{e}}_z$ is the unitary vector along z, $v \equiv \beta f l$ and

$$\mathcal{L}(v) \equiv \coth(v) - \frac{1}{v} \tag{3.37}$$

is the Langevin function. Analogously, the variances along the directions of the three axes are

$$\sigma_x^2 = \sigma_y^2 = l^2 g, \qquad \sigma_z^2 = l^2 a, \tag{3.38}$$

where (from now on, to ease notation we drop out the explicit v-dependence)

$$g \equiv \frac{\mathcal{L}}{v}, \qquad a \equiv 1 - 2g - \mathcal{L}^2. \tag{3.39}$$

Moreover, all the covariances involving different directions are zero, as expected from a random walk. Since $\boldsymbol{R_e}$ is the sum of independent identically-distributed variables, in the limit of large N we can apply the Central Limit Theorem and conclude that $\boldsymbol{R_e}$ is distributed according to a Gaussian. Moreover, from equation (3.36) we can argue that

$$\langle \boldsymbol{R_e} \rangle = N l \mathcal{L}(v) \hat{\boldsymbol{e}}_z, \tag{3.40}$$

while the variances along the three directions are given by $N\sigma_x^2 = N\sigma_y^2 = Nl^2 g$ and $N\sigma_z^2 = Nl^2 a$. The probability distribution $P_\nu(\boldsymbol{R_e})$ can thus be written as

$$P_\nu(\boldsymbol{R_e}) = P_\perp(R_x) P_\perp(R_y) P_z(R_z), \tag{3.41}$$

where

$$P_\perp(R_\alpha) = \frac{e^{-\frac{R_\alpha^2}{2Nl^2 g}}}{\sqrt{2\pi Nl^2 g}} \tag{3.42}$$

and

$$P(R_z) = \frac{e^{-\frac{(R_z - Nl\mathscr{L})^2}{2Nl^2 a}}}{\sqrt{2\pi Nl^2 a}}. \tag{3.43}$$

For $\nu \to 0$, we obtain $\mathscr{L} \to 0$, $g \to 1/3$, $a \to 1/3$, so that the result for unstretched chains provided by equation (3.2) is retrieved. In the opposite limit of an almost rod-like chain ($\nu \to \infty$), we find $g \simeq 1/\nu$, $a \simeq 1/\nu^2$, $i.e.$ the variances tend to zero, as expected. It is worth noting that the fluctuations along the direction of the force tend to zero faster than the ones in the transverse direction, in agreement with known asymptotic results on other models of polymer chains [17].

From the exact distribution of $\boldsymbol{R_e}$, several quantities of interest can be computed. Among these, the average squared end-to-end distance $\langle \boldsymbol{R_e^2} \rangle$ can be straightforwardly evaluated to be equal to

$$\langle \boldsymbol{R_e^2} \rangle = Nl^2(1 - \mathscr{L}^2) + N^2 l^2 \mathscr{L}^2. \tag{3.44}$$

Analogously, by considering equation (3.7) and proceeding as in the section 3.1.3, we can also easily compute the radius of gyration, which is equal to

$$\langle R_g^2 \rangle = \frac{1}{6} Nl^2(1 - \mathscr{L}^2) + \frac{1}{12} N^2 l^2 \mathscr{L}^2. \tag{3.45}$$

As expected, we find the unperturbed results $\langle \boldsymbol{R_e^2} \rangle = Nb^2$, $\langle R_g^2 \rangle = Nb^2/6$ when $\nu \to 0$, and the typical values of a rod $\langle \boldsymbol{R_e^2} \rangle = N^2 b^2$, $\langle R_g^2 \rangle = N^2 b^2/12$ in the limit of infinite forces. Remarkably, for intermediate forces the squared end-to-end distance and radius of gyration are exactly given by the average of these two limits, weighted according to the monotonical function $\mathscr{L}^2$. Moreover, for any fixed N, the rod-like term is the leading one for forces above the scale

$$\nu_c \sim 1/\sqrt{N}. \tag{3.46}$$

For $\nu < \nu_c$, the relative fluctuations of the end-to-end vector are larger than $\langle \boldsymbol{R_e} \rangle$, thus pointing towards the presence of important finite-size effects in this regime, in agreement with previous perturbative results [146].

3.2.2 Details of MC simulations

The main features of the shape of a stretched FJC were investigated by means of both analytical computations and MC simulations. The latter were performed extracting the orientation of each tangent vector directly from the distribution $p(\boldsymbol{r})$. More in detail, thanks to the cylindric symmetry of the problem, the azimuthal angle ϕ could be simply extracted uniformly in the range $[0,2\pi]$. As for the polar angle θ, a little workaround was needed in order to map its distribution onto uniform sampling. Let x be a random number uniformly distributed in the range $[0,1]$. By construction, the infinitesimal probability to find a number in the interval $[x,x+dx]$ is simply given by dx. As for the angle θ, by considering only the polar contribution to equation (3.34), we easily find for the distribution of its cosine

$$\tilde{p}(\cos\theta) = \frac{\nu}{2\sinh\nu}e^{\nu\cos\theta}. \tag{3.47}$$

Since the mapping has to preserve the infinitesimal probability of corresponding values of θ and x, the following condition has to be satisfied:

$$\tilde{p}(\cos\theta)d\cos\theta = dx. \tag{3.48}$$

Integrating both sides, we thus get

$$\frac{1}{2\sinh\nu}\left[e^{\nu\cos\theta} - e^{-\nu}\right] = x, \tag{3.49}$$

where the integration constant was fixed by imposing $\theta(x=0) = -\pi$. Therefore, for each step the polar angle θ was computed by inverting

$$\cos\theta = \frac{\log(e^{\nu} - 2x\sinh\nu)}{\nu}. \tag{3.50}$$

For each value of N and ν, the mean values of the several quantities considered in this chapter are obtained by averaging 10^4 different realizations. Statistical error is estimated by normalizing the standard deviation of the results and, when not shown, is always smaller than the size of symbols in the figures.

3.2.3 The shape of a stretched chain

In order to characterize the shape of a stretched chain, we evaluated the eigenvalues of the inertia tensor and the corresponding eigenvectors by means of Monte Carlo simulations, according to the algorithm presented above. In fig.3.1 we show a schematic picture recapitulating the evolution of the polymer as the tension is increased. Particularly, we sketch typical enveloping ellipsoids obtained for a chain of size $N = 200$ and for several values of ν. As the picture qualitatively shows, the effect of an external force on a chain is twofold. For low values of ν, the tension mostly affects the orientation of the enveloping ellipsoid [147], aligning it along its direction (bottom-left region in fig.3.1). In this regard, in fig.3.2 we show the av-
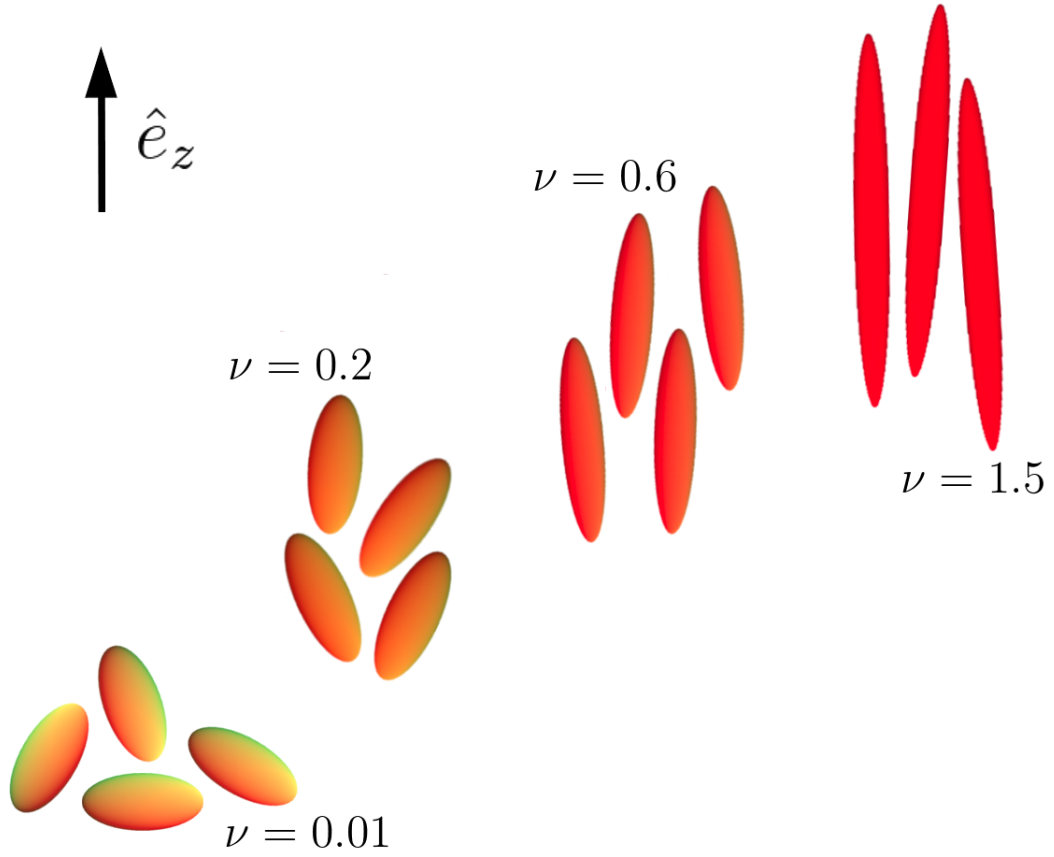
Figure 3.1 – Schematic illustration of the effect of an external force on a FJC. The arrow indicates the direction of the external force. All the ellipsoids shown have axes ratios corresponding to the average values of the eigenvalues for a chain made of $N = 200$ segments and for the values of ν reported in the figure. Particularly, we show the projection corresponding to the two largest eigenvalues. To ease visualization, the absolute size of the ellipsoids was chosen arbitrarily, thus relative sizes of ellipsoids drawn for different values of ν do not reflect the real values.

erage value of $\cos\psi_s$, where ψ_s is the angle between the applied force and the eigenvector corresponding to the eigenvalue λ_s. Starting from the typical value of random orientations (1/2), the cosine rapidly approaches 1 in the case of λ_1 (top) and 0 for λ_2 and λ_3 (bottom and inset), which corresponds to an ellipsoid with the principal axis oriented along the force. It is worth noting that the cosine approaches its large-ν value more fastly for larger sizes of the chain. After this dipole-like regime, the tension strongly deforms the polymer, thus leading to an increase in the anisotropy of its shape (top-right region in fig.3.1). However, we note that, due to the monotonicity of $\mathcal{L}^2$, R_g^2 increases monotonically from its unperturbed value to the rodlike limit (see equation (3.45)), thus a weak streching is present also at low forces. After the ellipsoid has been aligned, the largest eigenvalue λ_1 is expected to give the leading contribution to the squared radius of gyration R_g^2, thus a scaling dependence $\lambda_1 \sim N^2$ can
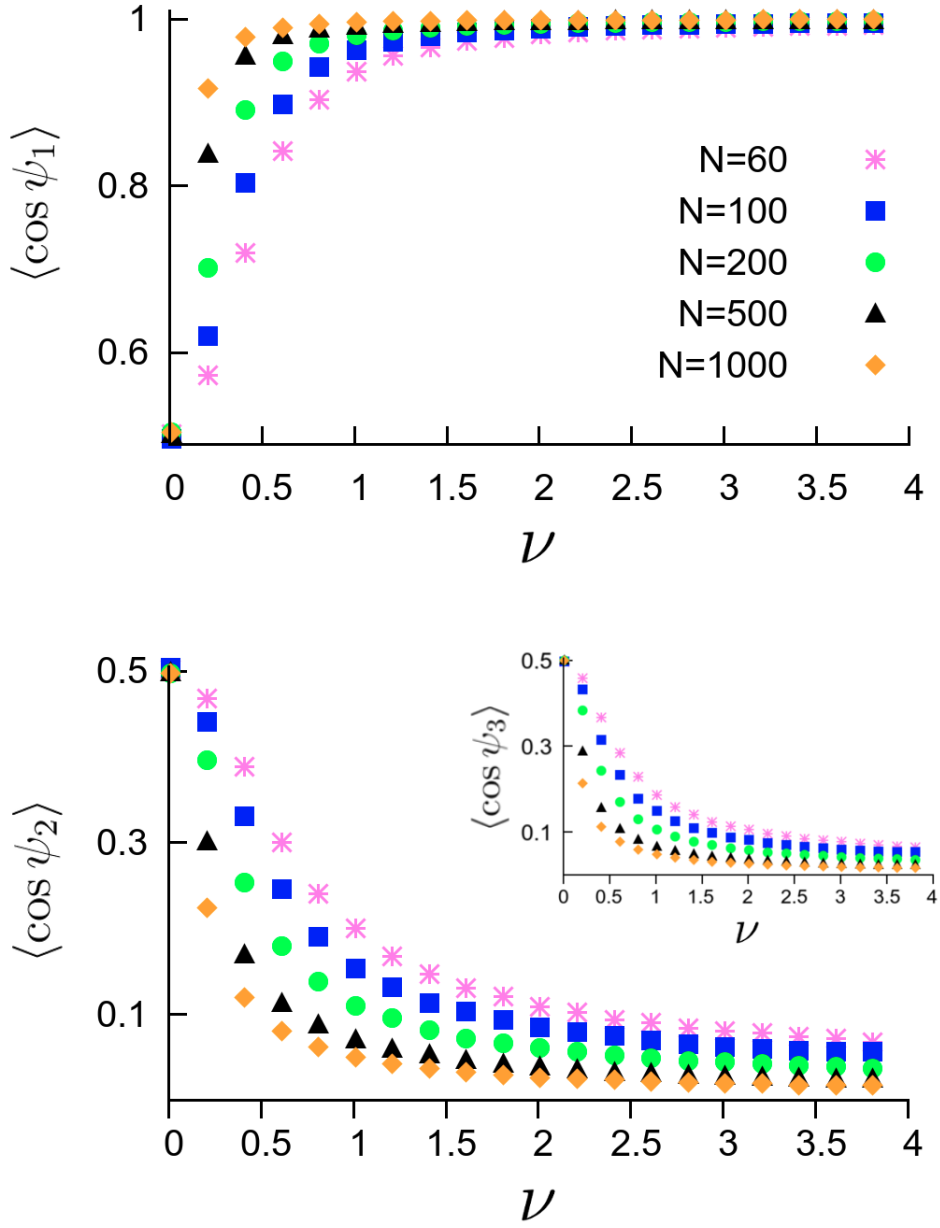
Figure 3.2 – Orientation of the ellipsoid enveloping a stretched FJC as a function of the adimensionalized force ν. In the plot we report for several values of N the average cosine of the angle ψ_s between the external force and the eigenvector corresponding to each eigenvalue for $s = 1$ (top), $s = 2$ (bottom) and $s = 3$ (inset). Adapted from [145].

already be anticipated. In fig.3.3 top we show $\lambda_1/N^2 l^2$ as a function of the adimensionalized force ν for several values of N. As expected, the various data sets collapse onto the same curve, though small but systematic displacements are observed for smaller chains. Since λ_1 gives the
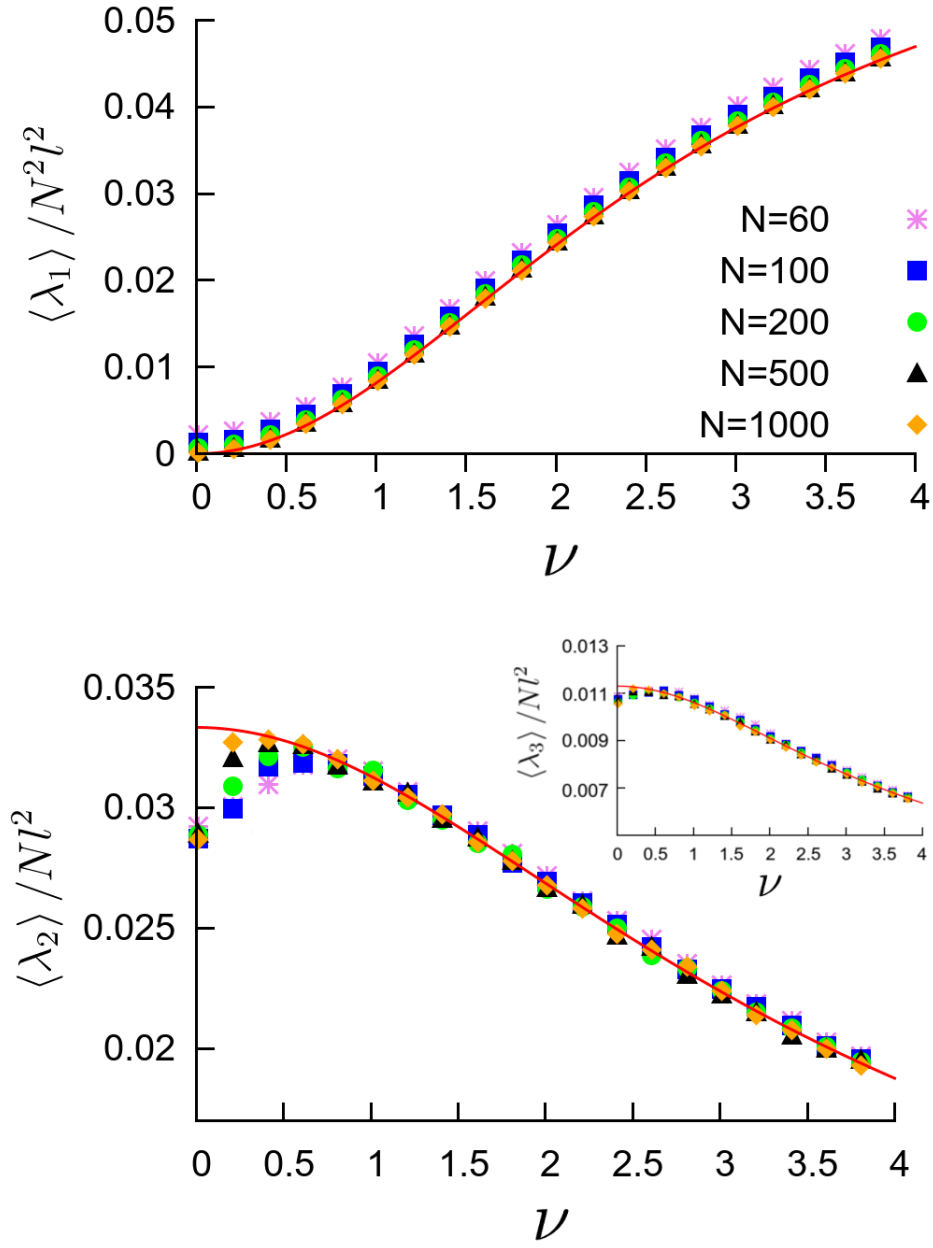
Figure 3.3 – Normalized eigenvalues $\lambda_1/N^2 b^2$ (top), λ_2/Nb^2 (bottom) and λ_3/Nb^2 (inset) as a function of the adimensionalized force ν, for several values of chain size N. The continuous curves are the corresponding fitting functions reported in equations (3.51), (3.52) and (3.53). Adapted from [145].

main contribution to the radius of gyration, in particular for large ν, the data are described

very well by the rodlike term in the formula for the radius of gyration (equation (3.45))

$$\frac{\lambda_1}{N^2 l^2} \simeq \frac{1}{12} \mathcal{L}^2, \tag{3.51}$$

which is reported as a continuous red line in fig.3.3 top. The two remaining eigenvalues show more interesting features. For large forces, they are expected to behave as the transverse contributions of R_g^2, which in turn are led by the fluctuations of the chain on the plane perpendicular to the force[2]. As a result, λ_2 and λ_3 are expected to behave as Ng (compare the formula for $P_\perp$, equation (3.42)). This ansatz is confirmed in fig.3.3, where we show that, for large enough ν, λ_2/Nl^2 (fig.3.3 bottom) and λ_3/Nl^2 (fig.3.3 inset) collapse onto universal curves well described by the functions

$$\frac{\lambda_2}{Nl^2} \simeq \frac{1}{10} g \tag{3.52}$$

and

$$\frac{\lambda_3}{Nl^2} \simeq \frac{1}{30} g, \tag{3.53}$$

which are plotted as red continuous lines in fig.3.3. Interestingly, the simulation data for λ_2
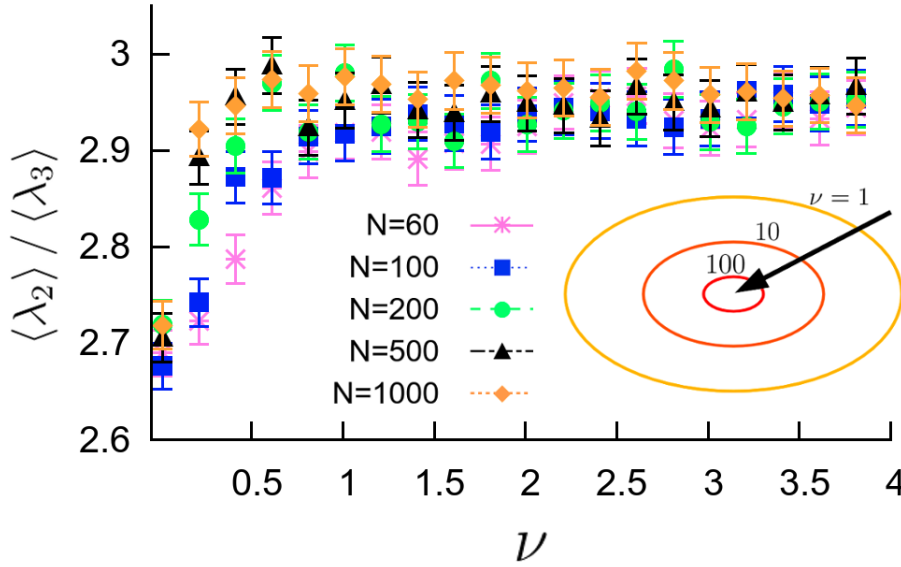

Figure 3.4 – Ratio between the average values of the two smallest eigenvalues. In the inset we report the shape of the transverse section of the chain for $\nu = 1, 10, 100$, where the axes are proportional to $\sqrt{\lambda_2}$ and $\sqrt{\lambda_3}$. The arrow indicates the direction of increasing ν.

---

[2]An explicit calculation performed by following the same approach as in section 3.1.3 shows that $\langle x_g^2 \rangle = \langle y_g^2 \rangle = Nl^2 g/6$, which has the same $N-$ and $\nu-$dependence of the fluctuations of the end-to-end vector in the transverse directions, as can be argued from equation (3.42).

and λ_3 show a non-monotonical behavior for small forces, which is not captured by equations (3.52) and (3.53). More in detail, starting from $v = 0$ they increase up to a maximum, after which they decrease according to the universal behavior. A comparison with the corresponding values of $\langle \cos \psi_s \rangle$ shows that the range of forces with increasing λ corresponds to a regime where the ellipsoid has still to align with the force. Therefore, an intuitive explanation of this phenomenon is that, due to the random orientation of the polymer (see bottom-left corner in fig.3.1), on average in this regime the force deforms the ellipsoid almost isotropically, thus leading to an increase of all the eigenvalues. In contrast, after a perfect alignment has been achieved (top-right corner in fig.3.1), only λ_1 keeps growing, while the two smaller eigenvalues shrink due to the smaller and smaller fluctuations in the directions perpendicular to the force. Notably, equations (3.52) and (3.53) suggest that at large forces the ratio between the two smaller eigenvalues is a constant. In this regard, in fig.3.4 we report the ratio $\langle \lambda_2 \rangle \langle / \lambda_3 \rangle$ as a function of the external force for several values of N. Starting from the unperturbed value 2.7 [143], the ratio increases with v until a plateau is reached. The stationary value is independent of N and is in good agreement with (though a bit lower than) the prediction obtained taking the ratio between equations (3.52) and (3.53), $\langle \lambda_2 \rangle / \langle \lambda_3 \rangle = 3$. We note that the increasing regime corresponds to the range of values of v where the alignment is taking place, as a direct comparison with fig.3.2 shows. The existence of a plateau outlines an intriguing picture of the evolution of a chain as the force is increased. After the ellipsoid has been aligned with the external force, as the latter increases the polymer becomes more and more elongated, while the size of the two smaller axes decreases. The constant ratio of the smallest eigenvalues shows that this shrinking is isotropic, thus the section maintains a universal shape independent of the force, as we sketch in the inset of fig.3.4.

We now focus on the computation of the asphericity. In this regard, we note that the approach introduced in section 3.1.3 can be straightforwardly extended to the case of a stretched chain. Indeed, equation (3.19) (and its analogue for $\langle (\mathrm{Tr}\,\mathcal{T})^2 \rangle$) is valid for any chain. Moreover, formulas like equation (3.22) and (3.23) show that, as long as the chain maintains its Markovian character, the key quantity to compute the moments of the inertia tensor is the squared average of the end-to-end vector. Therefore, in the present case we already have all the necessary tools to compute $\mathscr{A}$. The explicit calculation was performed in [142, 145] and is similar to what we did in section 3.1.3, thus we will not repeat it here. We limit ourselves to report the final results for the moments of $\mathcal{T}$ of interest, which are

$$\langle \mathcal{T}_{xx}^2 \rangle = \langle \mathcal{T}_{yy}^2 \rangle = \frac{g^2 N^2 l^4}{20}$$

$$\langle \mathcal{T}_{xy}^2 \rangle = \frac{g^2 N^2 l^4}{90}$$

$$\langle \mathcal{T}_{xx} \mathcal{T}_{yy} \rangle = \frac{g^2 N^2 l^4}{36}$$

$$\langle \mathcal{T}_{zz}^2 \rangle = \frac{N^2 l^4}{720} \left(5\mathscr{L}^4 N^2 + 44\mathscr{L}^2 N a + 36 a^2 \right)$$
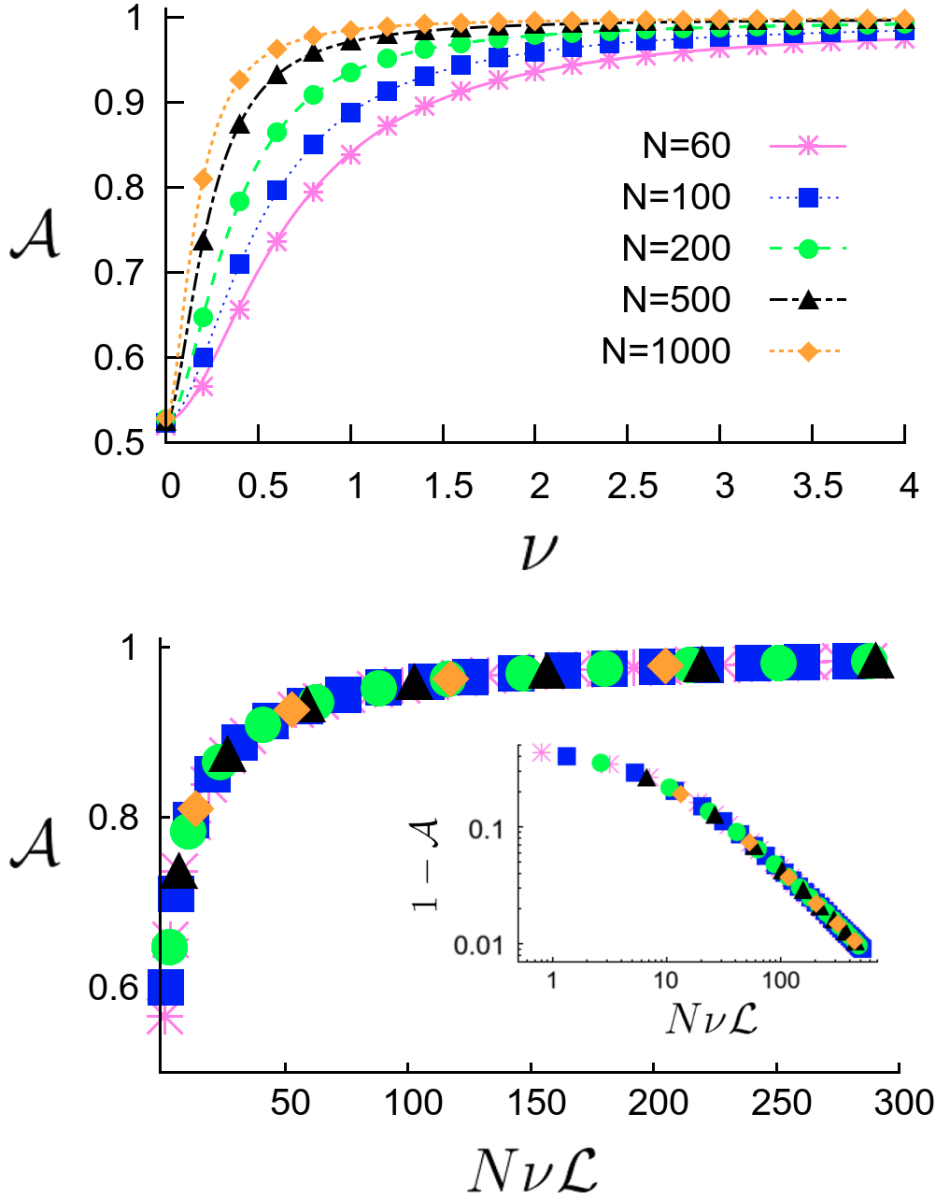
Figure 3.5 – Comparison between Monte Carlo data and exact formula (equation (3.54)) for the aspericity $\mathcal{A}$ as a function of the adimensionalized force v, for several values of N (top). If the data are plotted as a function of $Nv\mathcal{L}$, all the sets collapse onto a universal curve (bottom and inset). Reprinted from [145].

$$\langle \mathscr{T}_{xz}^2 \rangle = \langle \mathscr{T}_{yz}^2 \rangle = \frac{g N^2 l^4}{360} \left(4a + 3\mathscr{L}^2 N \right)$$

$$\langle \mathscr{T}_{xx}\mathscr{T}_{zz} \rangle = \langle \mathscr{T}_{yy}\mathscr{T}_{zz} \rangle = \frac{g N^2 l^4}{72} \left(2a + \mathscr{L}^2 N \right).$$

Note that some of the terms are equal because of the cylindric symmetry of the system. These terms can be used to compute both $\langle \mathrm{Tr}\left(\mathcal{T}^2\right)\rangle$ and $\langle(\mathrm{Tr}\,\mathcal{T})^2\rangle$, whose substitution in equation (3.18) finally gives

$$\mathcal{A} = 1 - \frac{72ag + 36g^2 + 24Ng\mathcal{L}^2}{36a^2 + 80ag + 112g^2 + 4\mathcal{L}^2(11a + 10g)N + 5N^2\mathcal{L}^4}.$$

(3.54)

In fig.3.5 top we plot Equation (3.54) as a function of v for different values of N. For comparison, we also show the average values of asphericity retrieved from our Monte Carlo simulations. Theoretical curves and simulation data perfectly agree with each other, providing a sanity check to the correctness of our calculations. Moreover, an interesting result comes from focusing on the coordinates x, y, *i.e.* on the plane orthogonal to the external force. Particularly, we can define the two-dimensional version of equation (3.18) as [143]

$$\mathcal{A}_{2d} \equiv \frac{2\langle \mathrm{Tr}\left(\mathcal{T}^2\right)\rangle}{\langle(\mathrm{Tr}\,\mathcal{T})^2\rangle} - 1$$

(3.55)

and compute the asphericity of the two-dimensional projection of the chain onto the xy plane. In the present case, we have

$$\langle \mathrm{Tr}\left(\mathcal{T}^2\right)\rangle = \langle\mathcal{T}_{xx}^2\rangle + \langle\mathcal{T}_{yy}^2\rangle + 2\langle\mathcal{T}_{xy}^2\rangle = \frac{11}{90}g^2N^2l^4$$

(3.56)

and

$$\langle(\mathrm{Tr}\,\mathcal{T})^2\rangle = \langle\mathcal{T}_{xx}^2\rangle + \langle\mathcal{T}_{yy}^2\rangle + 2\langle\mathcal{T}_{xx}\mathcal{T}_{yy}\rangle = \frac{7}{45}g^2N^2l^4.$$

(3.57)

Substitution in equation (3.55) finally gives

$$\mathcal{A}_{2d} = \frac{4}{7} \simeq 0.57.$$

(3.58)

Noticeably, the asphericity in the plane xy is independent of N and v, and has exactly the same value as the asphericity of an unperturbed two-dimensional FJC [143]. This result comes as no surprise, since it is due to the independence of the three directions of a random walk, and suggests that the conformations on the xy plane are the same as an unperturbed FJC where, in order to take into account the overall shrinking of the chain on this plane, the length of a segment has been properly renormalized. Nevertheless, at first sight it may seem that this result is not compatible with the ratio $\langle\lambda_2\rangle/\langle\lambda_3\rangle$ plotted in fig.3.4. Indeed, computing the two-dimensional asphericity corresponding to λ_2, λ_3 outlines a more spherical shape with $\mathcal{A}_{2d} \simeq 0.33$. These values are found for large v, long way after the alignment has taken place, thus the plane identified by considering the eigenvectors relative to λ_2, λ_3 should roughly correspond to the xy plane. How do we explain this apparent contradiction? Qualitatively, the key point is that, though the main axis of the ellipsoid becomes more and more aligned with the external force, the value of λ_1 increases with v, therefore its projection onto the xy plane could be comparable to the contributions coming from λ_2 and λ_3. If this is the

case, one cannot properly identify the projection of the FJC on the xy plane with the section perpendicular to the main axis of the ellipsoid, which would explain the apparent paradox reported above. In the next section, this observation will be addressed on a more quantitative level.

Though quite involved, equation (3.54) results very instructive when its limiting behavior is considered. Particularly, as long as $N\mathscr{L}^2 \gg 1$, the contributions coming from the size of the polymer dominate the behavior of the asphericity, which approaches unity as

$$\mathscr{A} \simeq 1 - \frac{24}{5Nv\mathscr{L}}. \tag{3.59}$$

For sufficiently large N, the previous condition can be rewritten as $Nv^2 \gg 1$, which points again to a force scale $v_c \sim 1/\sqrt{N}$ below which non-trivial finite-size effects are present (compare equation (3.46)). Furthermore, a Taylor expansion around $v = 0$ gives

$$\mathscr{A} \simeq \frac{10}{19} + \frac{25}{1083}Nv^2. \tag{3.60}$$

Since at low forces $Nv\mathscr{L} \sim Nv^2$, we can conclude that the asphericity depends on N and v by means of the combination $Nv\mathscr{L}$ in both the limiting cases, which suggests this to be the case, at the leading order, in the whole range of forces. As we show in fig.3.5 bottom, the data nicely collapse onto the same curve if $\mathscr{A}$ is plotted as a function of $Nv\mathscr{L}$, thus confirming our ansatz (a better insight on the large-force behavior is given in the inset of fig.3.5 bottom, where we plot $1 - \mathscr{A}$ in log-log scale).

3.2.4 A dipole approximation captures the orientational behavior of the FJC

The results reported in the previous section show that at the leading order the interplay between external force and polymer size affects the shape of the chain by means of the combination $Nv\mathscr{L}$. At this point, a natural question is whether also the orientation of the ellipsoid shows the same behavior. To answer this question, we considered again the average cosines reported in fig.3.2 and plotted them as a function of $Nv\mathscr{L}$. As we show in fig.3.6 and fig.3.7, this rescaling results again in a collapse of the data onto a universal curve in the whole range of explored values of the force. The alignment of the ellipsoid to the external force has a strong resemblance with the behavior of an electric dipole in the presence of an external field, with the important difference that in the former a 180° rotation results in the same physical state. It is therefore tempting to describe the observed curves for $\langle \cos\psi_s \rangle$ in terms of the dipole analog, with v as external field. In this regard, the dipole moment is epitomized by the elongation of the chain, thus we can interpret the factor $N\mathscr{L}$ as being proportional to the polarization response of the dipole: larger values of N result into a more responsive chain, though at large forces the dipole moment saturates to an asymptotic value. In other words, we assume the dipole to have a moment equal to $\alpha N\mathscr{L}$, where α is a proportionality constant.

Within this assumption, the interaction energy is equal to $\alpha Nv\mathscr{L} \cos\psi_1$, and the average
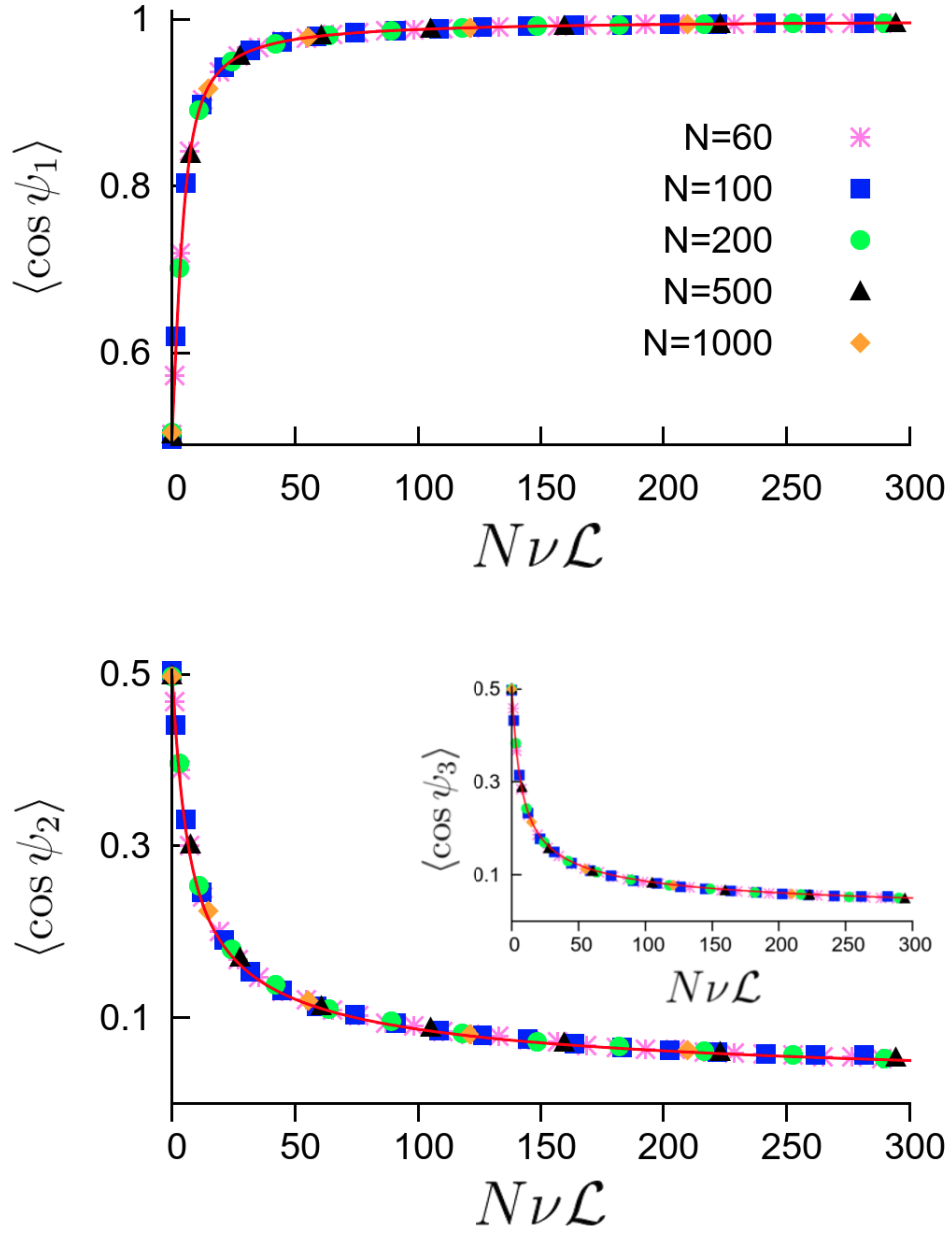
Figure 3.6 – Collapse of the average value of $\cos \psi_1$ (top, compare fig.3.2 top) when plotted as a function of $N\nu\mathcal{L}$. In the bottom panel the collapses of the other two angles are shown. The three continuous curves are obtained by globally fitting the data within the electric-dipole analog, with $\alpha \simeq 0.84$.

cosines are equal to

$$\langle\cos\psi_1\rangle = \frac{\int_0^1 \cos\psi\, e^{\alpha N\nu\mathcal{L}\cos\psi}\, d\cos\psi}{\int_0^1 e^{\alpha N\nu\mathcal{L}\cos\psi}\, d\cos\psi} = \frac{1}{1 - e^{-\alpha N\nu\mathcal{L}}} - \frac{1}{\alpha N\nu\mathcal{L}} \tag{3.61}$$
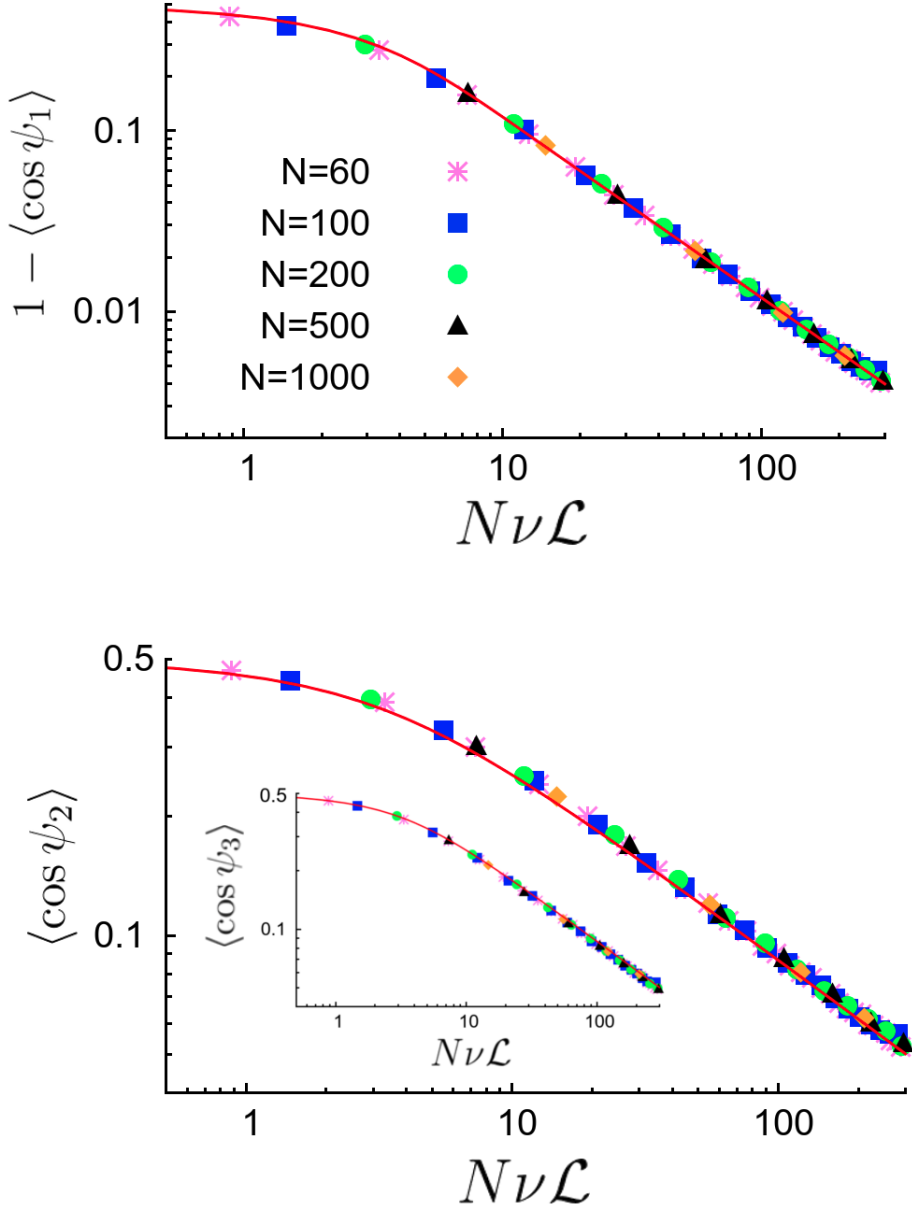
Figure 3.7 – Same data as in fig.3.7 plotted in logscale. Note that in the top panel we report $1 - \langle \cos \phi_1 \rangle$ in order to better analyze how the cosine approaches its asymptotic value.

and

$$\langle \cos \psi_{2,3} \rangle = \frac{\int_0^1 \cos \psi \, e^{\alpha N \nu \mathcal{L} \sin \psi} \, d\cos \psi}{\int_0^1 e^{\alpha N \nu \mathcal{L} \sin \psi} \, d\cos \psi} . \tag{3.62}$$

By tuning the parameter α, we performed a global fit of $\langle \cos \psi_s \rangle$, $s = 1, 2, 3$, for the set of data corresponding to $N = 1000$, and we found the continuous curves showed in fig.3.6 and fig.3.7.

The remarkable agreement between our ansatz and the simulation data shows that the dipole analog can capture even quantitatively the orientational behavior of the ellipsoid.
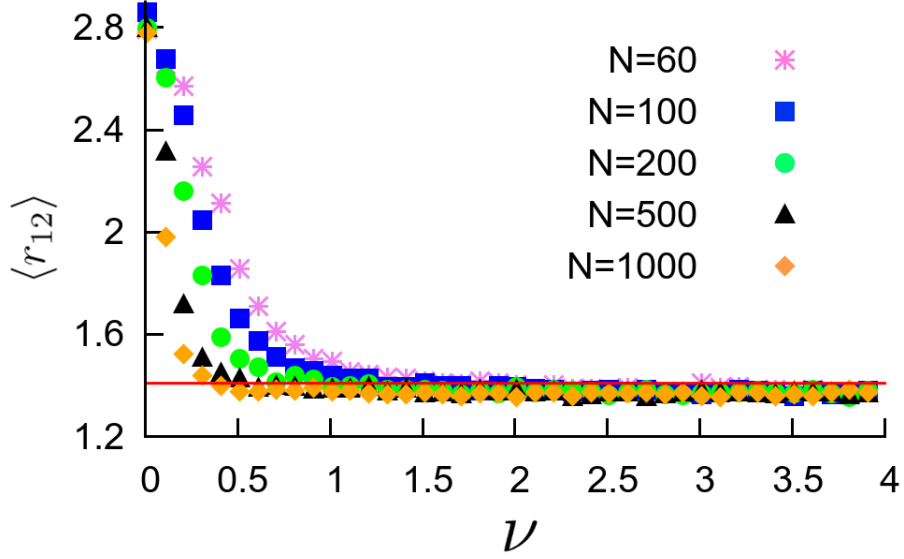

Figure 3.8 – Average value of r_{12} (equation (3.63)) as a function of ν for several values of N. The theoretical prediction for the asymptotic value (3.69) is shown as a continuous red line.

As a final comment, we can now address quantitatively the apparent paradox mentioned in the previous section. As we already mentioned, in order to properly identify the projection of the ellipsoid onto the xy plane with the section transverse to the eigenvector corresponding to λ_1, the $x-$ and $y-$ component of the major axis have to be negligible when compared to the corresponding values in the case of the minor ones. On a quantitative base, we thus have to evaluate *e.g.* the ratio

$$r_{12} \equiv \frac{\sqrt{\lambda_1}\sin\psi_1}{\sqrt{\lambda_2}\sin\psi_2}. \tag{3.63}$$

In order for the two projections to match each other, the condition

$$r_{12} \ll 1 \tag{3.64}$$

needs to be satisfied. Our results enable us to evaluate the average value of each factor appearing in equation (3.63). At large forces, equation (3.61) can be written as

$$\langle\cos\psi_1\rangle \simeq 1 - \frac{1}{\alpha N\nu\mathscr{L}}, \tag{3.65}$$

which thus gives

$$\langle\sin\psi_1\rangle \simeq \sqrt{\frac{2}{\alpha N\nu\mathscr{L}}}. \tag{3.66}$$

As for $\langle\sin\psi_2\rangle$, we note that at large forces $\langle\cos\psi_2\rangle$ is very small, thus the largest contribution to the integrals appearing in equation (3.62) is expected to come from low values of $\cos\psi$. We can thus write

$$\langle\cos\psi_2\rangle = \frac{\int_0^1 \cos\psi\, e^{\alpha N v \mathscr{L}\sqrt{1-\cos^2\psi}}\, d\cos\psi}{\int_0^1 e^{\alpha N v \mathscr{L}\sqrt{1-\cos^2\psi}}\, d\cos\psi} \simeq \frac{\int_0^1 \cos\psi\, e^{-\frac{1}{2}\alpha N v \mathscr{L}\cos^2\psi}\, d\cos\psi}{\int_0^1 e^{-\frac{1}{2}\alpha N v \mathscr{L}\cos^2\psi}\, d\cos\psi} =$$

$$\simeq \frac{\int_0^\infty \cos\psi\, e^{-\frac{1}{2}\alpha N v \mathscr{L}\cos^2\psi}\, d\cos\psi}{\int_0^\infty e^{-\frac{1}{2}\alpha N v \mathscr{L}\cos^2\psi}\, d\cos\psi} = \sqrt{\frac{2}{\pi\alpha N v \mathscr{L}}}\,. \tag{3.67}$$

Therefore, we have

$$\langle\sin\psi_2\rangle \simeq 1 - \frac{1}{\pi\alpha N v \mathscr{L}} \simeq 1\,. \tag{3.68}$$

Substituting equations (3.51), (3.52), (3.66) and (3.68) into equation (3.63), we finally get

$$r_{12} \simeq \frac{\sqrt{\frac{N^2\mathscr{L}^2}{12}}\sqrt{\frac{2}{\alpha N v \mathscr{L}}}}{\sqrt{\frac{Ng}{10}}} = \sqrt{\frac{5}{3\alpha}} \simeq 1.4\,, \tag{3.69}$$

independently of N and, more importantly, of v. In fig.3.8 we report the results from our Monte Carlo simulations for $\langle r_{12}\rangle$ as a function of the external force for several values of chain size. As predicted by equation (3.69), at large forces this ratio saturates to a value independent of the size N. The theoretical prediction, which is plotted in fig.3.8 as a red continuous line, is in good agreement with the simulation data, though it slightly overestimates the actual values. In this regard, we note that our approach completely neglected the possible presence of correlations between the various factors appearing in equation (3.63), which are thus likely to be responsible for the small quantitative disagreement between theory and simulations. To our purposes, the important point is that $\langle r_{12}\rangle$ reaches a plateau for large forces, so that the condition given by equation (3.64) is not satisfied. Therefore, λ_1 plays a non-negligible role in determining the shape properties of the FJC projection onto the xy plane, which solves the apparent contradiction between the values of the two-dimensional asphericities computed in the previous section.

3.3 Conclusions

In this chapter, starting from the exact distribution of the end-to-end vector, we have characterized in detail the properties of a stretched FJC. Our results show that both the shape ($\mathscr{A}$) and the orientation ($\langle\cos\psi_s\rangle$) of the polymer are due to finite-size effects. In the case of infinite N, any non-zero value of the force would result in a rod-like chain aligned with the external force. This behavior is in line with the linear-response "entropic spring" result, according to which for small v the relative elongation of the polymer along the direction of the

force follows a Hooke-like formula $\boldsymbol{f} = 3k_B T \boldsymbol{R_e}/Nb^2$ [27], which in the limit of infinite chain length suggests the polymer to be infinitely soft. Here, we have also quantitatively addressed the corrections introduced by finite values of N, showing that at the leading order both shape and orientation depend on N and v only through their combination $Nv\mathscr{L}$. Moreover, for increasing forces the chain shrinks in the plane orthogonal to the main axis of the ellipsoid, and it does so by maintaining a universal shape which has different features than those of an unperturbed two-dimensional FJC. Though derived on a FJC, our results for small forces can be directly applied to a wide variety of models, such as *e.g.* the Wormlike Chain in the case of double-stranded DNA [17], provided that the contour length of the chain is much larger than its Kuhn length [27].

Concluding Remarks

In this thesis we faced three different problems involving biomacromolecules. Particularly, the first two chapters have dealt with the study of specific biological systems, while the third one was focused on a more theoretical topic. These works give some examples of how biology can provide problems of great interest to physicists, and in turn how the latter can help understanding molecular mechanisms.

For example, inspired by some experiments on βlg fibrils, in chapter one we developed a model for the twisting properties of amyloid fibrils. Apart from being successfull in providing a quantitative description of the experimental results, the model goes beyond it and outlines the existence of universal laws governing the twisting properties of amyloid fibrils, which was confirmed in the case of flat-ribbon geometries by testing the universal predictions on sets of data involving different kinds of fibrils and different experimental techniques. Even more interestingly, as we explicitly showed by considering hollow-tube packing arrangements, different geometries are expected to result into different universal laws, thus highlighting the presence of "universality classes" of amyloid fibrils. This work shows how "old-style" physics can give some very deep insights on biomacromolecules, thus providing a useful complementary tool for their study.

In this regard, the coarse-grained analysis performed in chapter two is also quite significative. Apart from highlighting in a quantitative framework the fundamental role played by Hsp70 in the import of protein into mitochondria, it provides an instructive example of how coarse-grained simulations can help in understanding biological problems. Indeed, based on a mechanism proposed in the literature, we could show that basic ingredients such as the local flexibility of unfolded proteins and the excluded volume of amino acids can already result in import free energies leading to quantitatively-significant thermodynamic forces and gain in import times. Therefore, simple descriptions like the one we used can be very useful to identify the main players in biological processes. Naturally, more faithful representations would be needed in order to precisely fix the quantitative details.

In contrast with the previous chapters, the results derived in chapter three are not focused on a specific biological system, but they are rather more of theoretical interest. As already mentioned above, they are nonetheless insightful for biomolecular systems, since biopolymers are stretched in many instances within the cell, and their spatial distribution is important in

quantitatively determining the way they interact with other molecules. Our work fills a gap in the literature on this matter and paves the way for more detailed studies. Following the typical trend of basic results in polymer theory, our results outline the presence of universal functions capable of describing the behavior of chains of different size when subject to external forces. These results could be tested by performing experiments combining fluorescence and single-molecule techniques, though at present we are not aware of any such works aimed at investigating the shape of stretched chains.

In conclusion, we hope that after reading this thesis Alberts and coworkers would share our feeling that, within the microuniverse of the systems studied here, we have satisfactorily responded to their call, making a small further step in our understanding of cellular biology.

Bibliography

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell.* Garland Science, New York, 5th edition, 2007.

[2] G. Wider and K. Wüthrich. NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Curr. Opin. Struct. Biol.*, 9:594, 1999.

[3] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326:289, 2009.

[4] R. Y. Tsien. The green fluorescent protein. *Annu. Rev. Biochem.*, 67:509, 1998.

[5] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef. Mammalian genes are transcribed with widely different bursting kinetics. *Science*, 332:472, 2011.

[6] I. Gopich and A. Szabo. *Theory of Single-Molecule FRET Efficiency Histograms.* John Wiley and Sons, New York, 2011.

[7] L. Stryer. Fluorescence energy transfer as a spectroscopic ruler. *Annu. Rev. Biochem.*, 47:819, 1978.

[8] B. Cappella and G. Dietler. Force-distance curves by atomic force microscopy. *Surface Science Reports*, 34:1, 1999.

[9] T. E. Fisher, A. F. Oberhauser, M. Carrion-Vazquez, P. E. Marszalek, and J. M. Fernandez. The study of protein mechanics with the atomic force microscope. *Trends Biochem Sci.*, 24:379, 1999.

[10] J. R. Moffitt, Y. R. Chemla, S. B. Smith, and C. Bustamante. Recent advances in optical tweezers. *Annu. Rev. Biochem.*, 77:205, 2008.

[11] F. Ritort. Single-molecule experiments in biological physics: methods and applications. *J. Phys.: Condens. Matter*, 18:R531, 2006.

Bibliography

[12] S. B. Smith, L. Finzi, and C. Bustamante. Direct mechanical measurements of the elasticity of single dna molecules by using magnetic beads. *Science*, 258:1122, 1992.

[13] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. Reversible unfolding of individual titin immunoglobulin domains by afm. *Science*, 276:1109, 1997.

[14] M. B. Borgia, A. Borgia, R. B. Best, A. Steward, D. Nettels, B. Wunderlich, B. Schuler, and J. Clarke. Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature*, 474:662, 2011.

[15] A. Y. Grosberg and A. R. Khokhlov. *Statistical Physics of Macromolecules*. AIP Press, 1994.

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

[17] J. F. Marko and E. D. Siggia. Stretching DNA. *Macromolecules*, 28:8759, 1995.

[18] H. Hofmann, A. Soranno, A. Borgia, K. Gast, D. Nettels, and B. Schuler. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single molecule spectroscopy. *Proc. Nat. Academ. Sci.*, 109:16155, 2012.

[19] M. Karplus and J. A. McCammon. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.*, 9:646, 2002.

[20] ISI Web of Knowledge. http://www.webofknowledge.com.

[21] E. Khurana, M. Dal Peraro, R. DeVane, S. Vemparala, W. F. DeGrado, and M. L. Klein. Molecular dynamics calculations suggest a conduction mechanism for the M2 proton channel from influenza A virus. *Proc. Nat. Academ. Sci.*, 106:1069, 2009.

[22] D. Frenkel and B. Smit. *Understanding Molecular Simulations*. Academic Press, London, 2nd edition, 2002.

[23] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D*, 66:12, 2010.

[24] C. E. Lawrence, S. F. Altschul, M. S. Boguski, J. S. Liu, A. F. Neuwald, and J. C. Wootton. Detecting subtle sequence signals: A Gibbs sampling strategy for multiple alignment. *Science*, 262:208, 1993.

[25] J. P. Huelsenbeck and F. Ronquist. Mrbayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17:754, 2001.

[26] W. Liebermeister, T. A. Rapoport, and R. Heinrich. Ratcheting in Post-translational Protein Translocation: A Mathematical Model. *J. Mol. Biol.*, 305:643, 2001.

[27] M. Rubinstein and R. H. Colby. *Polymer Physics*. Oxford University Press, Oxford, 2003.

[28] T. E. Creighton. *Proteins: Structures and Molecular Properties - 2nd Ed.* W. H. Freeman and Company, New York, 1996.

[29] R. Phillips, J. Kondev, and J. Theriot. *Physical Biology of the Cell.* Garland Science, New York, 2008.

[30] C. Tanford. Protein denaturation: Part c. theoretical models for the mechanism of denaturation. *Adv. Protein Chem.*, 24:1, 1970.

[31] J. A. Schellman. Fifty years of solvent denaturation. *Biophys. Chem.*, 96:91, 2002.

[32] F. Chiti and C. M. Dobson. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.*, 75:333–366, 2006.

[33] V. N. Uversky and A. L. Fink. *Protein Misfolding, Aggregation and Conformational Diseases - Vols. 1,2.* Springer, Singapore, 2006.

[34] B. Caughey and P. T. Lansbury. Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.*, 26:267–298, 2006.

[35] C. M. Dobson. Protein folding and misfolding. *Nature*, 426:884–890, 2003.

[36] F. Chiti and C. M. Dobson. Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.*, 5:15–22, 2009.

[37] M. Sunde, L. C. Serpell, M. Bartlam, P. E. Fraser, M. B. Pepys, and C. C. F. Blake. Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *Journal of Molecular Biology*, 273(3):729 – 739, 1997.

[38] C. M. Dobson. Protein misfolding, evolution and disease. *Trends Biochem. Sci.*, 24:329, 1999.

[39] C. M. Dobson. The structural basis of protein folding and its links with human disease. *Philos. Trans. R. Soc. Lond B Biol. Sci.*, 356:133, 2001.

[40] C. M. Dobson. Getting out of shape—protein misfolding diseases. *Nature*, 418:729, 2002.

[41] V. N. Uversky and A. L. Fink. Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim. Biophys. Acta*, 1698:131, 2004.

[42] A. Fernandez and H. A. Scheraga. Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. *Proc. Natl. Acad. Sci. USA*, 100:113, 2003.

[43] J. Adamcik, J.-M. Jung, J. Flakowski, P. De Los Rios, G. Dietler, and R. Mezzenga. Understanding amyloid aggregation by statistical analysis of atomic force microscopy images. *Nat. Nanotechnol.*, 5:423, 2010.

[44] J. Adamcik and R. Mezzenga. Adjustable twisting periodic pitch of amyloid fibrils. *Soft matter*, 7:5437–5443, 2011.

[45] F. G. Pearce, S. H. Mackintosh, and J. A. Gerrard. Formation of amyloid-like fibrils by ovalbumin and related proteins under conditions relevant to food processing. *J. Agric. Food Chem.*, 55:318, 2007.

[46] G. M. Kavanagh, A. H. Clark, and S. B. Ross-Murphy. Heat-induced gelation of globular proteins: part 3. molecular studies on low ph β-lactoglobulin gels. *Int. J. Biol. Macromol.*, 28:41, 2000.

[47] S. G. Zhang. Fabrication of novel biomaterials through molecular self-assembly. *Nat. Biotechnol.*, 21:1171, 2003.

[48] I. Cherny and E. Gazit. Amyloids: Not only pathological agents but also ordered nano-materials. *Angew. Chem.*, 47:4062, 2008.

[49] E. Gazit. Self-assembled peptide nanostructures: the design of molecular building blocks and their technological utilization. *Chem. Soc. Rev.*, 36:1263, 2007.

[50] J. F. Smith, T. P. J. Knowles, C. M. Dobson, C. E. MacPhee, and M. E. Welland. Characterization of the nanoscale properties of individual amyloid fibrils. *Proc. Nat. Academ. Sci.*, 103:15806, 2006.

[51] T. P. J. Knowles, A. W. Fitzpatrick, S. Meehan, H. R. Mott, M. Vendruscolo, C. M. Dobson, and M. E. Welland. Role of intermolecular forces in defining material properties of protein nanofibrils. *Science*, 318:1900, 2007.

[52] T. P. J. Knowles, T. W. Oppenheim, A. K. Buell, D. Y. Chirgadze, and M. E. Welland. Nanostructured films from hierarchical self-assembly of amyloidogenic proteins. *Nat. Nanotechnol.*, 5:204, 2010.

[53] T. P. J. Knowles and M. J. Buehler. Nanomechanics of functional and pathological amyloid materials. *Nat. Nanotechnol.*, 6:469, 2011.

[54] J. L. Jiménez, E. J. Nettleton, M. Bouchard, C. V. Robinson, C. M. Dobson, and H. R. Saibil. The protofilament structure of insulin amyloid fibrils. *Proc. Nat. Academ. Sci.*, 99:9196, 2002.

[55] A. T. Petkova, Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio, and R. Tycko. A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. USA*, 99:16742, 2002.

[56] J. D. Harper, S. S. Wong, C. M. Lieber, and P. T. Lansbury. Assembly of Aβ amyloid protofibrils: An in vitro model for a possible early event in Alzheimer's disease. *Biochemistry*, 38:8972, 1999.

[57] A. W. P. Fitzpatrick, G. T. Debelouchina, M. J. Bayro, D. K. Clare, M. A. Caporini, V. S. Bajaj, C. P. Jaroniec, L. Wang, V. Ladizhansky, S. A. Müller, C. E. MacPhee, C. A. Waudby, H. R. Mott, A. De Simone, T. P. J. Knowles, H. R. Saibil, M. Vendruscolo, E. V. Orlova, R. G. Griffin, and C. M. Dobson. Atomic structure and hierarchical assembly of a cross-β amyloid fibril. *Proc. Nat. Academ. Sci.*, 110:5468, 2013.

[58] A. T. Petkova, R. D. Leapman, Z. Guo, W.-M. Yau, M. P. Mattson, and R. Tycko. Self-propagating, molecular-level polymorphism in Alzheimer's β-amyloid fibrils. *Science*, 307:262, 2005.

[59] A. K. Paravastu, R. D. Leapman, W. M. Yau, and R. Tycko. Molecular structural basis for polymorphism in Alzheimer's β-amyloid fibrils. *Proc. Natl. Acad. Sci. USA*, 105:18349, 2008.

[60] M. Fändrich, M. Schmidt, and N. Grigorieff. Recent progress in understanding Alzheimer's β-amyloid structures. *Trends Biochem. Sci.*, 36:338, 2011.

[61] X. Periole, A. Rampioni, M. Vendruscolo, and A. E. Mark. Factors that affect the degree of twist in β-sheet structures: A molecular dynamics simulation study of a cross-β filament of the GNNQQNY peptide. *J. Phys. Chem. B*, 113:1728, 2009.

[62] A. Aggeli, I. A. Nyrkova, M. Bell, R. Harding, L. Carrick, T. C. B. McLeish, A. N. Semenov, and N. Boden. Hierarchical self-assembly of chiral rod-like molecules as a model for peptide β-sheet tapes, ribbons, fibrils, and fibers. *Proc. Nat. Academ. Sci.*, 98:11857, 2001.

[63] T. P. J. Knowles, A. De Simone, A. W. Fitzpatrick, A. Baldwin, S. Meehan, L. Rajah, M. Vendruscolo, M. E. Welland, C. M. Dobson, and E. M. Terentjev. Twisting transition between crystalline and fibrillar phases of aggregated peptides. *Phys. Rev. Lett.*, 109:158101, 2012.

[64] I. A. Nyrkova, A. N. Semenov, A. Aggeli, and N. Boden. Fibril stability in solutions of twisted-sheet peptides: a new kind of micellization in chiral systems. *Eur. Phys. J. B*, 17:481, 2000.

[65] S. Bolisetty, J. Adamcik, and R. Mezzenga. Snapshots of fibrillation and aggregation kinetics in multistranded amyloid β-lactoglobulin fibrils. *Soft Matter*, 7:493, 2011.

[66] J. Adamcik and R. Mezzenga. Proteins fibrils from a polymer physics perspective. *Macromolecules*, 45:1137, 2012.

[67] S. Assenza, J. Adamcik, R. Mezzenga, and P. De Los Rios. Universal behavior in the mesoscale properties of amyloid fibrils. *Phys. Rev. Lett.*, 113:268103, 2014.

[68] J. M. Jung and R. Mezzenga. Liquid crystalline phase behavior of protein fibers in water: experiments versus theory. *Langmuir*, 26:504, 2010.

[69] L. D. Landau, E. M. Lifshitz, A. M. Kosevich, and L. P. Pitaevskii. *Theory of Elasticity*. Elsevier, Amsterdam, 3 edition, 1986.

Bibliography

[70] W. D. Pilkey. *Analysis and Design of Elastic Beams.* John Wiley and Sons, New York, 2002.

[71] J. N. Israelachvili. *Intermolecular and Surface Forces - 3rd Ed.* Academic Press, Burlington - MA, 2011.

[72] F. Chiti, P. Webster, N. Taddei, A. Clark, M. Stefani, G. Ramponi, and C. M. Dobson. Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proc. Natl. Acad. Sci. USA,* 96:3590, 1999.

[73] L. Nielsen, R. Khurana, A. Coats, S. Frokjaer, J. Brange, S. Vyas, V. N. Uversky, and A. L. Fink. Effect of environmental factors on the kinetics of insulin fibril formation: Elucidation of the molecular mechanism. *Biochemistry,* 40:6036, 2001.

[74] M. Silberberg. *Principles of General Chemistry.* McGraw-Hill, New York, 2006.

[75] C. Lara, J. Adamcik, S. Jordens, and R. Mezzenga. General self-assembly mechanism converting hydrolyzed globular proteins into giant multistranded amyloid ribbons. *Biomacromolecules,* 12:1868–1875, 2011.

[76] C. S. Goldsbury, G. J. Cooper, K. N. Goldie, S. A. Müller, E. L. Saafi, W. T. Gruijters, M. P. Misur, A. Engel, U. Aebi, and J. Kistler. Polymorphic fibrillar assembly of human amylin. *J. Struct. Biol.,* 119:17–27, 1997.

[77] G. S. Manning. Limiting laws and counterion condensation in polyelectrolyte solutions I. Colligative properties. *J. Chem. Phys.,* 51:924, 1969.

[78] E. R. P. Zuiderweg, E. B. Bertelsen, A. Rousaki, M. P. Mayer, J. E. Gestwicki, and A. Ahmad. Allostery in the Hsp70 Chaperone Proteins. *Top. Curr. Chem.,* 328:99, 2013.

[79] M. P. Mayer and B. Bukau. Hsp70 chaperones: cellular functions and molecular mechanism. *Cell. Mol. Life Sci.,* 62:670, 2005.

[80] P. Goloubinoff and P. De Los Rios. The mechanism of Hsp70 chaperones: (entropic) pulling the models together. *TRENDS in Biochem. Sci.,* 32:372, 2007.

[81] L. E. Hightower. Heat-shock, stress proteins, chaperones, and proteotoxicity. *Cell,* 66:191–197, 1991.

[82] F. U. Hartl and M. Hayer-Hartl. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science,* 295:1852–1858, 2002.

[83] S. Diamant, A. P. Ben-Zvi, B. Bukau, and P. Goloubinoff. Size-dependent disaggregation of stable protein aggregates by the DnaK chaperone machinery. *J. Biol. Chem.,* 14:21107, 2000.

[84] B. Bukau, E. Deuerling, and C. Pfund E. A. Craig. Getting newly synthesized proteins into shape. *Cell,* 101:119–122, 2000.

[85] C. Gaiddon, M. Lokshin, J. Ahn, T. Zhang, and C. Prives. A subset of tumor-derived mutant forms of p53 down-regulate p63 and p73 through a direct interaction with the p53 core domain. *Mol. Cell. Biol.*, 21:1874–1887, 2001.

[86] W. B. Pratt and D. O. Toft. Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. *Exp. Biol. Med. (Maywood)*, 228:111–133, 2003.

[87] T. Böcking, F. Aguet, S. C. Harrison, and T. Kirchhausen. Single-molecule analysis of a molecular disassemblase reveals the mechanism of Hsc70-driven clathrin uncoating. *Nat. Struct. Mol. Biol.*, 18:295–301, 2011.

[88] W. Neupert and M. Brunner. The protein import motor of mitochondria. *Nat. Rev. Mol. Cell. Biol.*, 3:555, 2002.

[89] K. E. Matlack, B. Misselwitz, K. Plath, and T. A. Rapoport. BiP acts as a molecular ratchet during posttranslational transport of prepro-alpha factor across the ER membrane. *Cell*, 97:553, 1999.

[90] L. Liu, R. T. McNeilage, L. X. Shi, and S. M. Theg. ATP requirement for chloroplast protein import is set by the Km for ATP hydrolysis of stromal Hsp70 in Physcomitrella patens. *The Plant Cell*, 113:121822, 2014.

[91] B. Liu, Y. Han, and S. Qian. Cotranslational Response to Proteotoxic Stress by Elongation Pausing of Ribosomes. *Mol. Cell*, 49:453, 2013.

[92] R. Shalgi, A. J. Hurt, I. Krykbaeva, M. Taipale, S. Lindquist, and C. B. Burge. Widespread Regulation of Translation by Elongation Pausing in Heat Shock. *Mol. Cell*, 49:439, 2013.

[93] E. B. Bertelsen, L. Chang, J. E. Gestwicki, and E. R. Zuiderweg. Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate. *Proc. Nat. Academ. Sci.*, 106:8471, 2009.

[94] R. Kityk, J. Kopp, I. Sinning, and M. P. Mayer. Structure and dynamics of the ATP-bound open conformation of Hsp70 chaperones. *Mol. Cell*, 48:863–874, 2012.

[95] R. S. Gupta. Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.*, 62:1435–1491, 1998.

[96] S. Rüdiger, L. Germeroth, J. Schneider-Mergener, and B. Bukau. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *EMBO J.*, 16:1501.x, 1997.

[97] D. Schmid, A. Baici, H. Gehring, and P. Christen. Kinetics of molecular chaperone action. *Science*, 263:971, 1994.

[98] M. P. Mayer. Hsp70 chaperone dynamics and molecular mechanism. *TRENDS in Biochem. Sci.*, 38:507–514, 2013.

[99] P. De Los Rios and A. Barducci. Hsp70 chaperones are non-equilibrium machines that achieve ultra-affinity by energy consumption. *eLife*, 3:e02218, 2014.

[100] J. Jayakumar, R. T. Smolenski, C. C. Gray, A. T. Goodwin, K. Kalsi, M. Amrani, and M.H. Yacoub. Influence of heat stress on myocardial metabolism and functional recovery after cardioplegic arrest: a P-31 NMR study. *Eur. J. Cardiothorac. Surg.*, 13:467–474, 1998.

[101] M. Revington, Y. Zhang, G. N. Yip, A. V. Kurochkin, and E. R. Zuiderweg. NMR investigations of allosteric processes in a two-domain Thermus thermophilus Hsp70 molecular chaperone. *J. Mol. Biol.*, 349:163–183, 2005.

[102] H. Theyssen, H. P. Schuster, L. Packschies, B. Bukau, and J. Reinstein. The second step of ATP binding to DnaK induces peptide release. *J. Mol. Biol.*, 263:657–670, 1996.

[103] D. L. Montgomery, R. I. Morimoto, and L. M. Gierasch. Mutations in the substrate binding domain of the Escherichia coli 70 kDa molecular chaperone, DnaK, which alter substrate affinity or interdomain coupling. *J. Mol. Biol.*, 286:915–932, 1999.

[104] R. T. Youker and J. L. Brodsky. *Regulation of Hsp70 function: hsp40 co-chaperones and nucleotide exchange factors.* Springer, New York, 2007.

[105] H. H. Kampinga and E. A. Craig. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat. Rev. Mol. Cell Biol.*, 11:579–592, 2010.

[106] D. Brehmer, C. Gassler, W. Rist, M. P. Mayer, and B. Bukau. Influence of GrpE on DnaK-substrate interactions. *J. Biol. Chem.*, 279:27957–27964, 2004.

[107] P. De Los Rios, A. Ben-Zvi, O. Slutsky, A. Azem, and P. Goloubinoff. Hsp70 chaperones accelerate protein translocation and the unfolding of stable protein aggregates by entropic pulling. *Proc. Nat. Academ. Sci.*, 103:6166, 2006.

[108] B. S. Glick. Can Hsp70 Proteins Act as Force-Generating Motors? *Cell*, 80:11, 1995.

[109] N. M. Toan, D. Marenduzzo, and C. Micheletti. Inferring the diameter of a biopolymer from its stretching response. *Biophys. J.*, 89:80–86, 2005.

[110] W. Sung and P. J. Park. Polymer Translocation through a Pore in a Membrane. *Phys. Rev. Lett.*, 77:783, 1996.

[111] L. Brocchieri and S. Karlin. Protein length in eukaryotic and prokaryotic proteomes. *Nucl. Acids Res.*, 33:3390, 2005.

[112] S. Assenza, P. De Los Rios, and A. Barducci. Quantifying the role of chaperones in protein translocation by computational modelling. *Front. Mol. Biosci.*, 2:8, 2015.

[113] A. Ghavami, E. Van der Giessen, and P. R. Onck. Coarse-Grained Potentials for Local Interactions in Unfolded Proteins. *J. Chem. Theory Comput.*, 9:432, 2013.

[114] X. Zhu, X. Zhao, W. F. Burkholder, A. Gragerov, C. M. Ogata, M. E. Gottesman, and W. A. Hendrickson. Structural analysis of substrate binding by the molecular chaperone DnaK. *Science*, 272:1606, 1996.

[115] S. Plimpton. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.*, 117:1, 1995.

[116] R. L. Burden and J. D. Faires. *Numerical Analysis*. Brooks Cole, Belmont, CA, 7th edition, 2000.

[117] T. Elston. Models of post-translational protein translocation. *Biophys. J.*, 79:2235, 2000.

[118] T. Elston. The Brownian ratchet and power stroke models for posttranslational protein translocation into the endoplasmic reticulum. *Biophys. J.*, 82:1239, 2002.

[119] A. Soranno, B. Buchli, D. Nettels, R. R. Cheng, S. Müller-Späth, S. H. Pfeil, A. Hoffmann, E. A. Lipman, D. E. Makarov, and B. Schuler. Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proc. Nat. Academ. Sci.*, 109:17800, 2012.

[120] J. H. Lim, F. Martin, B. Guiard, N. Pfanner, and W. Voos. The mitochondrial Hsp70-dependent import system actively unfolds preproteins and shortens the lag phase of translocation. *The EMBO Journal*, 20:941, 2001.

[121] Q. Liu, P. D'Silva, W. Walter, J. Marszalek, and E. A. Craig. Regulated Cycling of Mitochondrial Hsp70 at the Protein Import Channel. *Science*, 300:139, 2003.

[122] M. P. Mayer, H. Schröder, S. Rüdiger, K. Paal, T. Laufen, and B. Bukau. Multistep mechanism of substrate binding determines chaperone activity of Hsp70. *Nat. Struct. Biol.*, 7:586, 2000.

[123] M. Eilers and G. Schatz. Binding of a specific ligand inhibits import of a purified precursor protein into mitochondria. *Nature*, 322:228, 1986.

[124] A. J. Wilcox, J. Choy, C. Bustamante, and A. Matouschek. Effect of protein structure on mitochondrial import. *Proc. Nat. Academ. Sci.*, 102:15435, 2005.

[125] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, Amsterdam, 1992.

[126] G. Tiana, L. Sutto, and R. A. Broglia. Use of the Metropolis algorithm to simulate the dynamics of protein chains. *Physica A*, 380:241, 2007.

[127] E. Sanz and D. Marenduzzo. Dynamic Monte-Carlo versus Brownian dynamics: a comparison for self-diffusion and crystallization in colloidal fluids. *J. Chem. Phys.*, 132:194102, 2010.

Bibliography

[128] K. Ghosh and K. Dill. Cellular Proteomes Have Broad Distributions of Protein Stability. *Biophys. J.*, 99:3996, 2010.

[129] M. A. Scidmore, H. H. Okamura, and M. D. Rose. Genetic interactions between KAR2 and SEC63, encoding eukaryotic homologues of DnaK and DnaJ in the endoplasmic reticulum. *Mol. Biol. Cell*, 3:1145, 1993.

[130] R. I. Dima and D. Thirumalai. Asymmetry in the shapes of folded and denatured states of proteins. *J. Phys. Chem. B*, 108:6564–6570, 2004.

[131] C. E. Cordeiro, M. Molisana, and D. Thirumalai. Shape of confined polymer chains. *J. Phys. II France*, 7:433–447, 1997.

[132] D. J. Bonthuis, C. Meyer, D. Stein, and C. Dekker. Conformation and dynamics of DNA confined in slitlike nanofluidic channels. *Phys. Rev. Lett.*, 101:108303, 2008.

[133] C. Micheletti and E. Orlandini. Numerical study of linear and circular model DNA chains confined in a slit: metric and topological properties. *Macromolecules*, 45:2113–2121, 2012.

[134] G. Zifferer and W. Preusser. Monte carlo simulation studies of the size and shape of ring polymers. *Macromol. Theory Simul.*, 10:397–407, 2001.

[135] K. Alim and E. Frey. Shapes of semiflexible polymer rings. *Phys. Rev. Lett.*, 99:198102, 2007.

[136] E. J. Rawdon, J. C. Kern, M. Piatek, P. Plunkett, , A. Stasiak, and K. C. Millett. Effect of knotting on the shape of polymers. *Macromolecules*, 41:8281–8287, 2008.

[137] W. K. Lim and A. R. Denton. Polymer crowding and shape distributions in polymer-nanoparticle mixtures. *J. Chem. Phys.*, 141:114909, 2014.

[138] M. S. Cheung, D. Klimov, and D. Thirumalai. Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc. Nat. Academ. Sci.*, 102:4753–4758, 2005.

[139] D. Homouz, M. Perham, A. Samiotakis M. S. Cheung, and P. Wittung-Stafshede. Crowded, cell-like environment induces shape changes in aspherical protein. *Proc. Nat. Academ. Sci.*, 105:11754–11759, 2008.

[140] C. Bustamante, Z. Bryant, and S. B. Smith. Ten years of tension: single-molecule DNA mechanics. *Nature*, 421:423–427, 2003.

[141] W. Kuhn. Über die Gestalt fadenförmiger Moleküle in Lösungen. *Kolloid-Zeitschrift*, 68:2–15, 1934.

[142] A. S. Sassi. *Asymmetries in 3-dimensional unbranched polymer chains*. Master Thesis, EPFL, 2014.

[143] J. Rudnick and G. Gaspari. The aspherity of random walks. *J. Phys. A: Math. Gen.*, 19:L191–L193, 1986.

[144] J. Rudnick and G. Gaspari. *Elements of the Random Walk*. Cambridge University Press, Cambridge, 2004.

[145] A. S. Sassi, S. Assenza, and P. De Los Rios. The shape of a stretched polymer. *In preparation.*

[146] R. M. Neumann. On the precise meaning of extension in the interpretation of polymer-chain stretching experiments. *Biophys. J.*, 85:3418–3420, 2003.

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

Salvatore Assenza

Curriculum Vitae

EPFL SB ITP LBS
BSP 728
CH-1015 Lausanne
Switzerland
☎ *+41 21 69 30566*
✉ *salvatore.assenza@epfl.ch*

Personal Information

Born on | June 25th, 1986
Birthplace | Ragusa, Italy
Nationality | Italian

Current Position

Postdoctoral Fellow, *Laboratoire de Biophysique Statistique*, Ecole Polytechnique Fédérale de Lausanne, Switzerland (from April 2015)

Research Interests

○ Statistical Physics of polymers
○ Coarse-grained modeling of biomacromolecules
○ Stochastic processes in soft matter
○ Statistical Physics of Complex Networks

Publications

○ Quantifying the role of chaperones in protein translocation by computational modelling
S. Assenza, P. De Los Rios and A. Barducci
Front. Mol. Biosci. **2**:8 (2015)

○ Universal behaviour in the mesoscale properties of amyloid fibrils
S. Assenza, J. Adamcik, R. Mezzenga and P. De Los Rios
Phys. Rev. Lett. **113**,268103 (2014)

○ Emerging meso- and macro-scale structures from synchronization of adaptive networks
R. Gutierrez, A. Amann, S. Assenza, J. Gómez-Gardeñes, V. Latora and S. Boccaletti
Phys. Rev. Lett. **107**,234013 (2011)

○ Emergence of structural patterns out of synchronization in networks with competitive interactions
S. Assenza, R. Gutierrez, J. Gómez-Gardeñes, V. Latora and S. Boccaletti
Nature Sci. Rep. **1**,99 (2011)

○ Enhancement of cooperation in highly-clustered scale-free networks
S. Assenza, J. Gómez-Gardeñes and V. Latora
Phys. Rev. E **78**,017101 (2008)

Preprints & Research in Progress

○ The shape of a stretched chain
A. S. Sassi, S. Assenza and P. De Los Rios
in preparation
○ Polymer Simulations in Confined Geometries
D. Malinverni, S. Assenza and P. De Los Rios
in preparation
○ A novel approach to compute force-extension relations in polymer models
S. Assenza and P. De Los Rios
○ Thermodynamic efficiency of Hsp70 as unfoldase
S. Assenza, B. Schuler, P. De Los Rios and A. Barducci
○ Cooperativity in the pulling force of concatenated Hsp70 chaperones
M. Rizzi, S. Assenza, A. Barducci and P. De Los Rios

Oral Contributions

Contributed Talks in Conferences

Jul 2014　*Universality of Behaviour in the Mesoscale Properties of Amyloid Fibrils*, Annual Meeting of the Swiss Physical Society, Fribourg (Switzerland)

Jan 2014　*Exploring the Entropic Pulling by Molecular Dynamics*, International Conference on "25 years of chaperone research: Protein Folding, in and out of Anfinsen's closet", Arolla (Switzerland)

Feb 2012　*A Simple Model for Amyloid Fibril Twisting Structure*, Swiss Soft Days, 7th Workshop, Lausanne (Switzerland)

Other Oral Contributions

Feb 2014　*Exploring the Entropic Pulling by Molecular Dynamics*, IPSB Sandwich Seminar, Lausanne (Switzerland)

Feb 2014　*Universality of Behaviour in the Mesoscale Properties of Amyloid Fibrils*, Winter School on "Chromatin organization: from structure to biological function", Crans-Montana (Switzerland)

Jul 2012　*A simple model for amyloid fibrils twisting structure*, Summer School on "Polymers in soft and biological matter", Boulder (Colorado, USA)

Jun 2012　*Twisting structure of amyloid fibrils emerges as a compromise between competing forces*, "Diploma di Licenza Specialistica" defence, Scuola Superiore di Catania*, Catania (Italy)

Oct 2010　*Synchronization in adaptive networks with competitive interactions*, Master Degree defence, University of Catania (Italy)

Nov 2009　*Emergence of Synchronization in Nonlinearly Coupled Oscillators*, "Diploma di Licenza Triennale" defence, Scuola Superiore di Catania*, Catania (Italy)

Nov 2008	*Enhancement of Cooperation in Highly-Clustered Scale-Free Networks*, Bachelor Degree defence, University of Catania (Italy)
Oct 2008	*Enhancement of Cooperation in Highly-Clustered Scale-Free Networks*, Seminar at Ecole Polytechnique Fédérale de Lausanne (Switzerland)
Jul 2008	*Enhancement of Cooperation in Highly-Clustered Scale-Free Networks*, Summer School for Prospective Master Students at Scuola Normale Superiore di Pisa, S. Miniato, Italy

Posters

Conferences

Mar 2015	*Quantifying the role of chaperones in protein translocation by computational modelling*, EMBO Conference on "Mechanisms and regulation of protein translocation", Dubrovnik (Croatia)
Sep 2014	*Universality of Behaviour in the Mesoscale Properties of Amyloid Fibrils*, IBI workshop on "Bioengineering day - breaking down disciplinary barriers", Lausanne (Switzerland)
Sep 2012	*Twisting Structure of Amyloid Fibrils Emerges as a Compromise between Competing Forces*, Fall Meeting of the Swiss Chemical Society, Zürich (Switzerland)

Other poster presentations

Feb 2014	*Universality of Behaviour in the Mesoscale Properties of Amyloid Fibrils*, Winter School on "Chromatin Organization - from structure to biological function", Crans-Montana (Switzerland)
Jul 2013	*Universality of Behaviour in the Mesoscale Properties of Amyloid Fibrils*, Summer School on "Biomolecules and single molecules techniques", Miraflores (Spain)
Feb 2013	*Universality of Behaviour in the Mesoscale Properties of Amyloid Fibrils*, Winter School on "Simulating activated processes in Physics and Chemistry: theoretical foundations", Villars (Switzerland)
Jul 2012	*Twisting Structure of Amyloid Fibrils Emerges as a Compromise between Competing Forces*, Summer School on "Polymers in Soft and Biological Matter" Boulder (Colorado, USA)
Jan 2012	*A Simple Model for Amyloid Fibril Twisting Structure*, Winter School on "Cellular Biophysics : Molecules, membranes, and mechanics", Ascona (Switzerland)

Visiting

Nov 2012	National Institute for Medical Research, London, UK
Feb-Mar 2010	Universidad Rey Juan Carlos, Madrid, Spain
Jan 2010	Northeastern University, Boston, Massachusetts (USA)
Oct 2008	École Polytechnique Fédérale de Lausanne, Switzerland

Participation in Schools

| Feb 2014 | *Chromatin Organization - from structure to biological function*, Crans-Montana (Switzerland) |

Jul 2013	*Biomolecules and single molecules techniques*, Miraflores (Spain)
Feb 2013	*Simulating activated processes in Physics and Chemistry: theoretical foundations*, Villars (Switzerland)
Jul 2012	*Polymers in Soft and Biological Matter*, Boulder (Colorado, USA)
Jan 2012	*Cellular Biophysics : Molecules, membranes, and mechanics*, Ascona (Switzerland)

Education

2011
2015

PhD, Physics, *Ecole Polytechnique Fédérale de Lausanne*, Switzerland.

Dissertation *Polymer-Physics insights into biomolecular systems*

Supervisors Prof. Paolo De Los Rios and Dr. Alessandro Barducci

2009
2012

"Diploma di Licenza Specialistica", Physics, *Scuola Superiore di Catania**, Italy.

Final Grade 70/70 cum laude

Dissertation *Twisting Structure of Amyloid Fibrils Emerges as a Compromise between Competing Forces*

Supervisor Prof. Paolo De Los Rios

2008
2010

Master Degree in Physics, *University of Catania*, Italy.

Final Grade 110/110 cum laude

Dissertation *Synchronization in Adaptive Networks with Competitive Interactions*

Supervisors Prof. Vito Latora and Prof. Jesús Gómez-Gardeñes

2005
2009

"Diploma di Licenza Triennale", Physics, *Scuola Superiore di Catania**, Italy.

Final Grade 70/70 cum laude

Dissertation *Emergence of synchronization in nonlinearly coupled oscillators*

Supervisors Prof. Vito Latora and Prof. Jesús Gómez-Gardeñes

2005
2008

Bachelor Degree in Physics, *University of Catania*, Italy.

Final Grade 110/110 cum laude

Dissertation *Enhancement of cooperation in highly-clustered Scale-Free Networks*

Supervisors Prof. Vito Latora and Prof. Jesús Gómez-Gardeñes

Scholarships and Awards

2011	National award "Roberto Giordano" for the best Student in Theoretical Physics by University of Catania
2005-2010	Scholarship within Scuola Superiore di Catania*
2009	Scolarship for Bachelor Degree transcripts by "Provincia di Ragusa"
2006	Scolarship for High School transcripts by "Provincia di Ragusa"
2005	Scolarship for High School transcripts by "Provincia di Ragusa"

*The "Scuola Superiore di Catania" (SSC) is an excellence center within University of Catania where, after a

competitive selection based on written and oral exams, a very restricted number of undergraduate students receive a scholarship for the whole duration of their studies fully covering living expenses and fees. Moreover, apart from the normal courses at University of Catania, SSC students attend additional courses aimed at boosting an early start of their research activity. For further information see www.scuolasuperiorecatania.it

Teaching Activity

2014 **Teaching Assistant**, *Mechanique Analytique (Analytical Mechanics, course held in French)*, BSc course, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

2013 2014 **Co-advisor for Master Project**, *Asymmetries in three-dimensional unbranched polymer chains*, A. S. Sassi, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

2012 2014 **Teaching Assistant**, *Statistical Physics of Biomacromolecules*, MSc Course, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

2012 2014 **Teaching Assistant**, *Biothermodynamics*, BSc Course, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

2011 2013 **Tutor**, *Travaux Pratique 4 (pre-thesis project for MSc students)*, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

2010 2011 **Tutor**, *Calculus, Analytical Mechanics and Mathematics for Economics for BSc students*, Collegio d'Aragona, Catania, Italy.

Skills

Languages

- Italian, mothertongue
- English, fluent
- Spanish, fluent
- French, fluent
- German, basic

Applications

Mathematica, Matlab, Office, Gimp, Latex, GNUplot, VMD

Programming and Scripting

C/C++, LAMMPS, Bash, SLURM, awk, Python

References

Prof. Paolo De Los Rios
École Polytechnique Fédérale de Lausanne
Laboratoire de Biophysique Statistique
CH-1015 Lausanne, Switzerland
✉ paolo.delosrios@epfl.ch
☎ +41 21 69 30510

Dr. Alessandro Barducci
École Polytechnique Fédérale de Lausanne
Laboratoire de Biophysique Statistique
CH-1015 Lausanne, Switzerland
✉ alessandro.barducci@epfl.ch
☎ +41 21 69 30513

Prof. Vito Latora
Queen Mary, University of London
School of Mathematical Sciences
Mile End Road, London E1 4NS, United Kingdom
✉ latora@maths.qmul.ac.uk
☎ +44 020 7882 5199

Jesús Gómez-Gardeñes
University of Zaragoza
Dep. Fisica de Materia Condensada
C/ Pedro Cerbuna 12, E-50009 Zaragoza, Spain
✉ gardenes@gmail.com
☎ +34 976 76 1221