

Tecnologica

2013

MARCO BACCI

# Coarse-grained molecular dynamics and continuum models for the transport of protein molecules

P R E M I O T E S I D O T T O R A T O  
F I R E N Z E U N I V E R S I T Y P R E S S – U N I V E R S I T À D E G L I S T U D I D I F I R E N Z E



PREMIO TESI DI DOTTORATO

- 44 -

PREMIO TESI DI DOTTORATO  
Commissione giudicatrice, anno 2013

Luigi Lotti, *presidente della Commissione*

Tito Arecchi, *Area Scientifica*

Franco Cambi, *Area Umanistica*

Paolo Felli, *Area Tecnologica*

Michele Arcangelo Feo, *Area Umanistica*

Roberto Genesisio, *Area Tecnologica*

Luigi Lotti, *Area Scienze Sociali*

Mario Pio Marzocchi, *Area Scientifica*

Adolfo Pazzagli, *Area Biomedica*

Mario Giuseppe Rossi, *Area Umanistica*

Salvatore Ruggieri, *Area Biomedica*

Saulo Sirigatti, *Area Biomedica*

Piero Tani, *Area Scienze Sociali*

Fiorenzo Cesare Ugolini, *Area Tecnologica*

Vincenzo Varano, *Area Scienze Sociali*

Graziella Vescovini, *Area Umanistica*

Marco Bacci

**Coarse-grained molecular dynamics and  
continuum models for the transport of  
protein molecules**

Firenze University Press  
2014

Coarse-grained molecular dynamics and continuum models for the transport of protein molecules / Marco Bacci. – Firenze : Firenze University Press, 2014. (Premio Tesi di Dottorato; 44)

<http://digital.casalini.it/9788866556947>

ISBN 978-88-6655-693-0 (print)

ISBN 978-88-6655-694-7 (online)

#### *Peer Review Process*

All publications are submitted to an external refereeing process under the responsibility of the FUP Editorial Board and the Scientific Committees of the individual series. The works published in the FUP catalogue are evaluated and approved by the Editorial Board of the publishing house. For a more detailed description of the refereeing process we refer to the official documents published in the online catalogue of the FUP ([www.fupress.com](http://www.fupress.com)).

#### *Firenze University Press Editorial Board*

G. Nigro (Co-ordinator), M.T. Bartoli, M. Boddi, R. Casalbuoni, C. Ciappei, R. Del Punta, A. Dolfi, V. Fargion, S. Ferrone, M. Garzaniti, P. Guarnieri, A. Mariani, M. Marini, A. Novelli, M. Verga, A. Zorzi.

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0: <http://creativecommons.org/licenses/by/4.0/>)

**CC** 2014 Firenze University Press  
Università degli Studi di Firenze  
Firenze University Press  
Borgo Albizi, 28, 50122 Firenze, Italy  
[www.fupress.com](http://www.fupress.com)  
*Printed in Italy*

Ai miei zii, zie, cugini e cugine



*“Do not trust any existing theory. Start to think independently from as few intuitive concepts as possible to form your own conclusion.”*

René Descartes





# Table of contents

<b>Abstract</b>	<b>9</b>
<b>Acknowledgements</b>	<b>11</b>
<b>Chapter 1.</b>	
<b>Introduction</b>	<b>13</b>
1. Preamble	13
2. Translocation in biology	13
2.1. Protein molecules	14
3. Voltage-driven experiments and nanopore technology	15
4. Outline of the thesis	19
<b>Chapter 2.</b>	
<b>State of the Art</b>	<b>21</b>
1. Preamble	21
2. Experimental work: a selected example from the literature	21
3. Additional past work outline	24
4. Analytical models and simulations of protein translocation	28
4.1. Overview	28
4.2. Coarse-grained molecular dynamics reference studies	29
<b>Chapter 3.</b>	
<b>Molecular Dynamics Translocation Simulations</b>	<b>35</b>
1. Preamble	35
2. Three-dimensional models for MBP simulations	36
2.1. MBP numerical model: The Gō-like approach	37
2.2. Pore and pulling models	40
2.3. Langevin dynamics for the material points of the MBP	41
3. Methods for translocation and stretching simulations	42
3.1. Translocation simulations	42
3.2. Stretching simulations	43
4. Translocation and stretching numerical results	43

4.1. Denaturation characterization	43
4.2. Translocation dynamics	45
4.3. Residence time statistics	47
4.4. Stretching vs. translocation	51
5. Discussion	54
6. Remarks	58
<b>Chapter 4.</b>	
<b>1-Dimensional Continuum Models</b>	<b>61</b>
1. Preamble	61
2. Overview	61
2.1. The general framework	61
3. Models and methods	65
4. Results	67
5. Remarks	77
<b>Chapter 5.</b>	
<b>3-Dimensional Continuum Models</b>	<b>79</b>
1. Preamble	79
2. Introduction	79
3. Models and methods	81
3.1. Discrete scheme	81
3.2. Continuum model	81
3.3. Stretching simulations: methods	84
3.4. Translocation simulations: methods	85
4. Results	85
4.1. Toy simulations and related remarks	85
4.2. Three-dimensional atomic clusters	92
4.3. Additional remarks	98
4.4. Protein descriptor in MBP stretching simulations	104
4.5. MBP morphological descriptor translocation results	105
5. Additional continuum formulations	107
5.1. Varying the meaning of $\nu$	108
5.2. $z_s$ model 3 and 3*	111
5.3. Remarks	115
<b>References</b>	<b>119</b>

## Abstract

In the present thesis, coarse-grained molecular dynamics simulations and relevant continuum models are investigated. In the first part we present a computational analysis on the driven transport of Maltose Binding Protein (MBP) across nano-channels in the framework of coarse-grained modeling. The work is motivated by recent experiments on voltage driven transport of MBP across nanopores exploring the influence of denaturation on translocation pathways. Our simplified approach allows a statistical mechanics analysis of the transport phenomenology which is useful to investigate the influence of denaturation from a structural point of view. Numerical results are qualitatively in agreement with experimental data. They reproduce a rich phenomenology, from fast transport to long blockades (Fig. 1) and, to some extent, bumping events. Specifically, we identify and characterize short and long channel blockades, associated with the translocation of denatured and folded MBP structures, respectively. We show that long blockades are related to stall events where MBP undergoes specific and reproducible structural rearrangements. To clarify the origin of the stalls, the

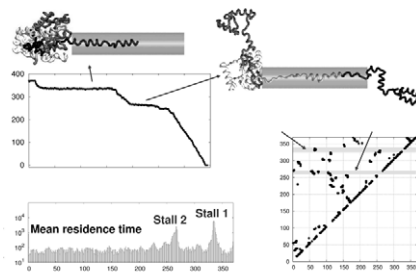


Figure 1: Stall snapshots during translocation of folded MBP, statistics of residence time and native contact map.

*stick-and-slip* translocation is compared with mechanical unfolding pathways obtained via steered molecular dynamics. This comparison clearly shows that translocation pathway differs significantly from free-space unfolding dynamics and strongly suggests that stalling events are preferentially determined by MBP regions with higher density of long-range native interactions. This result might constitute a possible criterion to predict *a-priori* statistical features of protein translocation from structural analysis.

Customarily, protein dynamics is summarized in low-dimensional models where only the evolution of few order parameters is considered as an appropriate description of the complete dynamics of the atomic cluster constituting the molecule. We show that long stall events and main translocation features are essentially maintained in the 1D description of the process selected in the present work, where the evolution of an appropriate collective variable and the relevant free energy landscape are considered. Moreover, we show that both free energy barriers and translocation bottlenecks are strongly associated with the structural properties of the folded proteins. In particular, free-energy ramps systematically correspond to the regions along the backbone chain with greater density of long-range native contacts with the untranslocated portion of the protein itself. These areas are also responsible for the stalls during nanopore transport. Also, this knowledge allow us to decompose the free energy trends into the different contributions of *internal energy* and *entropic effect*, pointing out the dominant terms. Here, results relevant to two different proteins and to a mutant one are considered and discussed.

In the second part of the thesis, we enrich the standard 1D continuum view by proposing a continuum model for the description of the dynamics of isolated macromolecules. We adopt a second-rank tensor as a descriptor of the macromolecular shape and identify the action governing its dynamics by means of an identification procedure from a discrete scheme, based on power equivalence and Cauchy-Born rule. We compare molecular dynamics stretching simulations with the continuum model by starting from discrete toy schemes, going on with increasing complexity, and ending with the analysis of the Ubiquitin protein. The results indicate limitations in the approach in case of unconstrained molecular dynamics while they show appropriateness for driven dynamics (Fig. 2). Also MBP is considered by addressing primarily translocation simulations, a particular constrained dynamics where the selected approach confirms its appropriate-

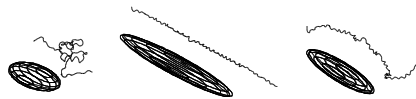


Figure 2: Snapshots of molecule Ubiquitin and relevant morphological descriptor during a mechanical pulling simulation.

ness.

Finally, to overcome the shortcomings stemming from possible lack of homogeneous deformation Cauchy-Born rule is based on, we investigate some alternative heuristic approaches, which appear adequate in describing the protein shape evolution also in unconstrained dynamics, such as free equilibrations.

## Acknowledgements

First and foremost, I wish to acknowledge sincerely my advisor, Prof. Paolo Maria Mariano. Motivations to my gratefulness are several, and start earlier than the research activity here presented, when just after the master degree, I asked him for a piece of advice to pursue the deep interest I have always felt for biomechanics. I can state, the PhD. days have been tough for me, nevertheless for him, supporting my work and intellectual growth from the very beginning and providing me with many opportunities, even beyond the mere scientific context. He has never lacked of attention and real passion toward our work and the academic world. In these last days his detailed revision to this manuscript has been crucial and tireless.

I also wish to express my gratitude to Prof. Roberto Genesio and Prof. Stefano Ruffo, for the support during the long journeys abroad and for the interest they have always shown in the research activity I have been involved in.

This research project would not have been possible without the support of many people. Profound thanks are directed to the research group in Rome, Dr. Fabio Cecconi, Prof. Carlo Massimo Casciola and Dr. Mauro Chinappi. I have performed a considerable part of the project in collaboration with them, who in turn have always considered me as a full member of the group, carrying out stimulating and challenging discussions, which can not be neglected neither from a scientific nor from a personal point of view. I will remember the experience in Rome as a wonderful period, where I started to focus on the research proposal in a stimulating and friendly environment, the so-called ‘aquarium’. Really, a special thank, again, to Mauro Chinappi, he taught me things I should have learnt before, with unflinching patience and competence.

I am grateful to Dr. Sandrasekaram Gnanakaran, Dr. Giovanni Bellesia and to the whole T6 in Los Alamos National Laboratory. Despite the short visit, I have been involved in meetings and discussions that considerably fostered the second part of the research project about continuum modeling. Also, feeling at home in such a far away country is not for granted and the friendly environment of the Lab, despite what one might wrongly believe, played a substantial rôle.

I would like to gratefully acknowledge Prof. Charles S. Peskin and the Courant Institute in New York University. By far the toughest days during the PhD., as it has to be in such a mathematical school. Thanks to Prof. Peskin I had the opportunity to present our scientific results to a selected group of scientists, who have not been short of comments and hints for future research

(finger crossed). However, above all, I have had the possibility to interact with a person who undertakes his job driven by passion, thus inspiring an otherwise impossible serenity and simplicity to all the ones who have the chance (and luck) to work with him. A big 'thank you' also to Giulio Trigila, Abba and Shoshana Leffler, Adam Stinchcombe and friends, colleagues and staff at NYU.

Mentioning the New York City days, I can not forget my girlfriend Jocelyn. That is another story, I know, and deserves another stage. Let me just say I am impressed and thankful to her determination and strength. 24/7 working schedule for both of us, and not a complaint. I received only sympathy, sharing, admiration and love. Many many thanks to Donna and Alex, Angie and Willie and, of course, Reyes and their families, who have welcome me as if I had always been there.

Thank you zio Gigi, always so eager to interact and work with me, another old-fashioned person that goes on for true passion.

My family, I have been far apart for some time, but it has always been like having you close by. I have always felt a deep, sincere and excited interest for what I have been on. Thank you, knowing your estimation for my commitment allowed me not to give up.

Moreover, I wish to thank my friends, a bearing column where to rest, take a break and dream.

Finally, I thank the personnel of University of Florence, and the Ministry of Public Education for the scholarship I have benefitted from.

# Chapter 1

## Introduction

### 1. Preamble

The present dissertation is supervised by Prof. Paolo Maria Mariano of University of Florence and Scuola Normale Superiore of Pisa.

It is developed in collaboration with Prof. Carlo Massimo Casciola and Dr. Mauro Chinappi of Sapienza University of Rome and Dr. Fabio Cecconi, Institute of Complex Systems, National Research Council of Rome.

The focus of the work is on the transport of proteins across nanopores, a process called translocation. The work is based on coarse-grained molecular dynamics simulations exploiting the so-called Gō-like model, [1, 2]. Equilibration, mechanical stretching and transport simulations are developed.

A standard 1D continuum approach based on the Langevin equation in the potential of the mean force relevant to a suitably-defined reaction coordinate is investigated. The main aim is to establish its suitability to retain and interpret the complex three-dimensional molecular dynamics phenomenology.

Non-conventional continuum modeling is also considered. In this perspective a second rank tensor is defined in order to summarize the three-dimensional shape evolution of clusters of atoms in space, in turn undergoing several different dynamics. The analysis is performed in light of the mechanics of complex materials [3].

### 2. Translocation in biology

Translocation, namely the passage of molecules through biological pores, is a fundamental process occurring in many natural phenomena, [4]. It involves the exchange of proteins, ions, energy sources, genetic information and any particle or aggregate that plays a role in cell functioning and living. In particular, biopolymer translocation across membranes is ubiquitous in biology and includes the passage of RNA strands inside nuclear pores [5], absorption of oligonucleotides on suitable sites [6], transport of proteins back and forth from the cells [7], mitochondrial import, protein degradation by ATP-dependent proteases and protein synthesis, [8, 9].



Cells are filled with all kinds of nanopores, committed to an overwhelming range of functions. Proteins hold an important role in transport processes across cell membranes. Indeed the latter are generally composed of phospholipid bilayers and the channels across them by transmembrane proteins. The transport can be either passive or active. The former one can take place by diffusion or by interposition of carriers to which the transported substances bonded. In the active case instead, transmembrane proteins use the Adenosine-Tri-Phosphate chemical energy to drive a molecule against its concentration gradient. A mechanism of selection, based on the interaction between signal recognition particles and relative receptors, allows the transmembrane proteins to detect the system to be transported [7].

As it will be described in Sec. 3, the discovery of the protein nature of transmembrane pores fostered the synthesis of conductive channels to perform *in vitro* translocation experiments. Although the first pioneering experimental work on a transmembrane biopore embedded in an artificial lipid bilayer, both housed in a microfluidic cell, dates back to the 90's [10], clarifying the physical and chemical phenomena involved in nanopore transport will remain a fertile field of research for the years to come, as it is only recently that scientists have developed suitable techniques to adequately tackle this issue.

Proteins are not only the basic constituent of transmembrane channels, but are also commonly transported between organelles and cells. Very often membrane apertures are not large enough to let the proteins enter in their native (folded) conformations. For example, the narrowest constriction in the proteosome organelle is only 13 Å in diameter. It is thus necessary for the proteins to unfold in advance or during their passage across the pore [11]. Such co-translocation unfolding rate is orders of magnitude faster than chemical or thermal denaturation, suggesting a different unfolding mechanism [9, 12]. A natural and appealing explanation of this gap is to suppose the process to be driven by electromechanical forces.

### 2.1. Protein molecules

Proteins are complex polymers composed of smaller elementary molecules called amino acids. The specific monomer sequence determines their three-dimensional structure, which is in turn responsible for the biological function of the protein itself. Proteins influence a large number of phenomena in life science. For instance, some of them are involved in the transmission of signals between cells and tissues (*hormones*), other ones support the immune system against pathogens (*antibodies*); moreover *enzymes* cause or foster the onset of chemical reactions otherwise impossible at physiological conditions. In addition, proteins are responsible for the regulation of gene expression by bonding with specific nucleic acid targets. Structural proteins are fundamental in converting chemical energy into mechanical work in muscles, besides constituting

part of the scaffold of tissues such as hair, nails, tendons and bones. Finally, particular proteins store and transport particles, control the passage of molecules across membranes and drive the electron current in plant photosynthesis. Proteins can be subdivided in two main categories:

1. globular proteins,
2. fibrous proteins.

The former are the majority, are soluble and their three-dimensional structure is almost spherical. The latter are insoluble and do not have a predetermined shape, although often are rod-like.

The conformation that allows a protein to comply with its biological task is said *native configuration*, or also *folded*, *natured* or *compact* state. There are several techniques to achieve protein denaturation. *Unfolded states* can be obtained, for example, either thermally, chemically or even mechanically, as in atomic force microscope experiments or by using optical tweezers.

In nature there exist 20 different standard amino acids that constitute, as already mentioned, the backbone chain of the protein molecules. Protein size ranges from 50 up to more than 3000 amino acids. However, the latter can bond one another forming also small peptides or polypeptides with no specific three-dimensional shape. Nineteen amino acids have the same general structure (depicted in Fig. 1.1) and differ only in the chemical composition of the side chain **R** (residue chain). The side chain of the 20<sup>th</sup> (prolyne) is standard, but it is bonded to a nitrogen atom instead than to a carbon atom. The latter is called *α-carbon atom*. Other carbon atoms along the side chain are termed *β, γ, δ, ...* accordingly to their position with respect to the first one. Very often amino acids are simply called *residues* in the biological parlance.

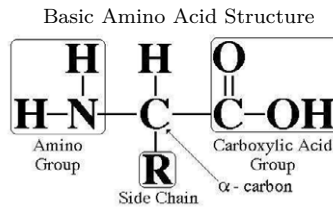


Figure 1.1: General representation of amino acids. Main components are highlighted and labeled.

### 3. Voltage-driven experiments and nanopore technology

The basic apparatus to perform voltage-driven translocation experiments is constituted by an electro-chemical micro-fluidic cell, as the one depicted in

Fig. 1.2 panel *a*, which represents the evolution of the Coulter resistive counter [13]. In such a device two chambers (*cis* and *trans*), which contain a buffer so-

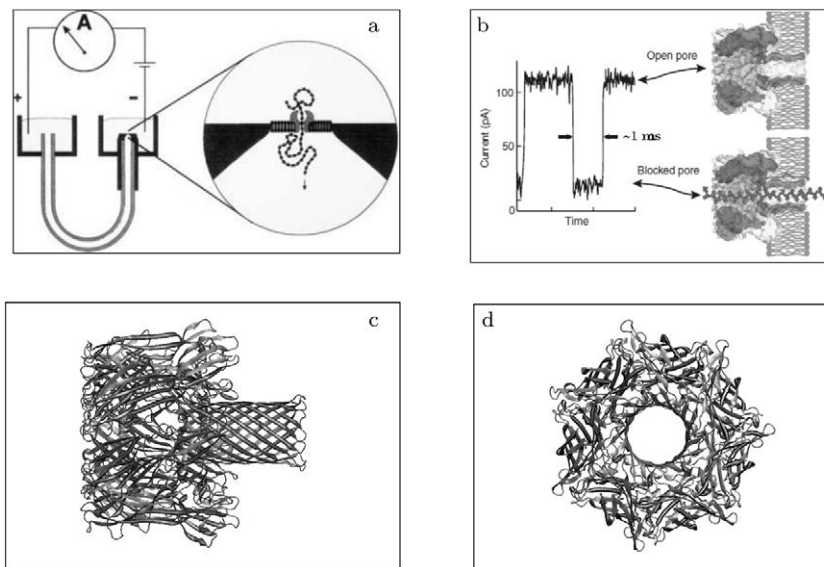


Figure 1.2: Snapshots of panels *d* and *c* are obtained using VMD software [14]. Panel *a*: Sketch of a microfluidic experimental apparatus for voltage-driven translocation experiments. A magnification of the sensing area of the device (nanopore) is also provided. Panel *b*: Standard ion current blockade event at the passage of a molecule across the channel. Panel *c*: Side view of the  $\alpha$ -hemolysin transmembrane protein. Panel *d*: Characteristic eptameric structure of the  $\alpha$ -hemolysin. This toxin is in fact composed of seven elementary proteins arranged in a circular fashion.

lution, are connected by a pipe that also hosts the sensing area, *i.e.* a membrane where a nanopore is inserted (single nanopore systems). An applied voltage, usually of the order of some  $mV$ , generates an ionic current through the channel that can be recorded by standard electro-physiological techniques like patch-clamp [15]. Once biopolymers (with a net charge at the pH of the solution) are introduced (on the *cis*-side), they cross the pore driven by thermal fluctuations (in the approaching to the pore) and by the voltage (across the pore itself, where the majority of the voltage drop takes place) and flow into the opposite chamber (*trans*-side). The passage temporarily clogs the channel and provokes a detectable ion current drop, Fig. 1.2 panel *b*, which strongly depends on the chemical and physical properties of the molecule that occupies the pore. For this reason, nanopore systems can work as efficient devices to operate at the molecular level in order to characterize several different biological components [16]. Nanopore-based technology is believed a promising resource offering powerful tools for detection [17], manipulation and sequencing of macromolecules [18],

although meaningful data analysis still remains challenging [19]. Moreover, the duration of the current drop is a direct measure of the translocation time, which is in turn used to characterize a single transport event, along with the intensity of the blockage itself. Clusters of events, usually coupled with changes in external conditions such as temperature or salt concentration, can as well provide information about the composition of the solution.

In voltage-driven translocation experiments, a widely exploited transmembrane protein is the well-known  $\alpha$ -hemolysin ( $\alpha$ HL), a  $\beta$ -barrel toxin expressed by the very common bacterium *Staphylococcus Aureus* [10].  $\alpha$ HL is very stable *in-vitro* and spontaneously migrates into lipid bilayers, considerably simplifying the building up of the experimental apparatus. Therefore it has been employed in several cases both for DNA/RNA strands and for protein translocation [10, 16, 20–24]. Its crystallographic structure was determined with a resolution of 1.9 Å [25] and a schematic view is given in panels *c* and *d* of Fig. 1.2.

Another biological nanopore that worth to be mentioned for similar applications is the Aerolysin channel [26], recently used by Pastoriza-Gallego *et al.* to study the translocation of a macromolecule called Maltose Binding protein [27, 28].

Other pioneering experimental studies concerned detecting polypeptides not only in the pore lumen but also in the aqueous phase. This analysis has been achieved for the first time by Movileanu and co-workers [29, 30], who were able to reveal the presence of a particular protein in the proximity of the biopore by attaching a flexible linker within the large cavity of the  $\alpha$ -hemolysin pore with a specific ligand at its end. Inspired by this work, Kong and Muthukumar performed Langevin molecular dynamics simulations and Poisson-Nerst-Planck calculations to model the fluctuations of the linker inside the vestibule [31], elucidating the dynamics of the single-molecule captures in the aqueous phase. Even if several other experiments worth to be referred, it is evident that the fundamental analyses just mentioned open the way to theoretical modeling and numerical investigation of an enormous amount of phenomena. Moreover, as Movileanu delineates [24], nanopores represent single-molecule probes with several possible applications. They can be employed to reveal many features of important biomolecules like folding state, backbone flexibility, mechanical stability, binding affinity and charge distribution, in addition to the possibility of revealing the biophysical rules that govern the transport through transmembrane channels. However, the main idea that fostered the development of voltage-driven translocation experiments, since the very first application, has always been the one to identify and fast-sequence the molecule that is clogging the throat section of the channel. Although the original approach concerned only DNA/RNA strands, protein sequencing is currently deemed even more interesting than DNA characterization. Indeed, whereas several fast and inexpensive techniques to perform DNA read-out already exist, the achievement of information on protein structures and amino acid order through standard experimental approaches such as x-

ray crystallography or nucleic magnetic resonance is still challenging and time-consuming. However protein sequencing is even more complex than DNA for several reasons. The latter in fact is linear and homogeneously charged, differently from a structured polymer. Also, the ability to discern between different elementary molecules has to be superior for protein characterization. Indeed DNA or RNA are constituted by only five different types of oligonucleotides while the basic constituents of protein molecules are the 20 standard amino acids.

Several groups have already investigated the interactions of small polypeptides, including  $\alpha$ -helical [32–34] and  $\beta$ -barrel [35] hairpin polypeptides, or even whole macromolecules [23, 36, 37] with the  $\alpha$ HL or anthrax pores. A crucial step toward single-molecule differentiation and sequencing has been recently carried out by Talaga and Li [38]. They have distinguished from each other different proteins in solution by modeling the characteristic ion current drop across a solid state nanopore through a simple Ohmic term, multiplied by a corrective factor that takes into account the relative geometry of the protein and the channel. This approach will be probably delved deeper in the near future to take into account other physical and chemical aspects of the molecules and of the whole system.

As already stated, first experiences were carried out by exploiting biological pores but, as mentioned in the last example, recent advances in technology have opened new possibilities toward the use of artificial solid state nanopores.

A great advantage of protein biopores is that they can be chemically engineered with advanced molecular biology techniques, such as mutagenesis [39]. Using this approach a wide variety of  $\alpha$ -hemolysin biosensors were developed, especially to probe binding affinities [30, 37, 40] and to reach discrimination at single-base resolution on isolated strands of DNA [41, 42]. Moreover, the geometry of a biopores is well-known and is highly reproducible, that is, once it has been determined by using, for example, x-ray crystallography, it can be assumed to be the same for all the biochannels of that particular type. Finally, toxin pores like  $\alpha$ HL or anthrax [23, 36, 43] embed spontaneously in lipid bilayer membranes. However they exhibit a number of disadvantages too. Typically the pores, but mainly the membranes that host them, can become unstable if changes are produced in the external environment in terms of temperature, pH or denaturant concentration (used to unfold the proteins in solution). Also, it is very difficult to imagine portable devices which sensing area is made of biological pores, due to instability of, above all, the lipid bilayer. In addition, their fixed size can also represent a limit, because geometry and dimensions cannot be macroscopically changed to address specific issues.

Solid state nanopores present some advantages over their counterparts, such as increased reliability and lifetime, control over surface properties and pore geometry (diameter, length and shape), compatibility with existing semiconductor and microfluidic fabrication techniques with thus potential integration into ultra-high throughput devices [44, 45]. In fact, nanopore-based DNA analysis

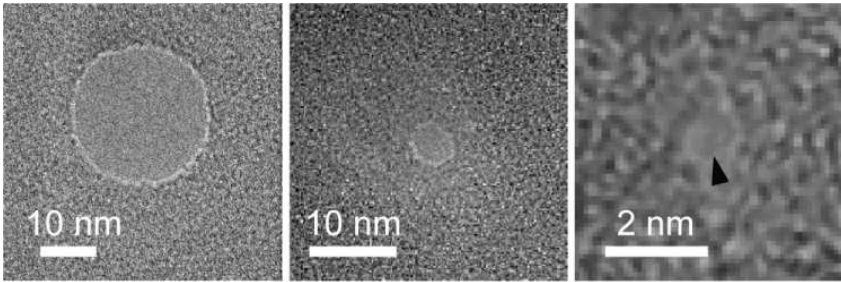


Figure 1.3: Transmission electron microscope images of a solid state nanopore. Adapted from [44].

has the potential to carry out a range of laboratory and medical experiments even faster than current methods, also reducing the cost (fast DNA sequencing is an application of nanopore based sensing that took about 25 years to transfer the early insights into a commercial device, announced by the end of 2012 [46]). Figure 1.3 shows TEM images of silicon nanopores. The first synthetic biosensor was developed by Charles Martin's group in 2005 and consisted of a single conical gold nanotube [47]. Since then, several teams have pursued this route, demonstrating the capability of solid state nanopores to perform single-molecule detection and to probe several characteristics of the proteins [38, 48–52].

Finally, as speculated by Branton *et al.* [53], the possible embedding of a biological pore into a semiconductor membrane represents a fascinating opportunity that should be carefully considered and evaluated.

#### 4. Outline of the thesis

The thesis is organized as it follows. In chapter 2 an overview on the state of the art of voltage-driven translocation experiments on proteins is provided, along with a summary of numerical investigations concerning coarse grained molecular dynamics. Specifically, the work by Oukhaled *et. al* [23] is described, since it is the experimental analysis that has motivated a substantial part of the numerical simulations performed during the present research activity. Also, numerical analyses [54, 55] are outlined for can be deemed the reference works on the topic available at the beginning of the thesis. Chapter 3 shows molecular dynamics results on protein stretching and translocation [56]. Chapter 4 is dedicated to standard one-dimensional modeling of the translocation process, in terms of driven diffusion of a proper reaction coordinate in the associated potential of the mean force [57]. Main theoretical aspects are also considered along with the exposition of the numerical results obtained. Finally, the last chapter contains the three-dimensional continuum analysis performed to enrich the just mentioned standard view. The analysis is performed in light of the mechanics

## Coarse-grained molecular dynamics and continuum models

of complex materials [3] and it is in particular inspired by two previous works [58, 59].

# Chapter 2

## State of the Art

### 1. Preamble

This chapter principally reviews the main experimental work that constitutes the baseline of our research activity, to which the whole section 2 is dedicated.

However, a more general description of both experimental and numerical investigations is provided in section 3, where also a popular coarse grained model somewhat similar to the Gō model is outlined. It is worth to anticipate that the Gō model, sec. 2.1, constitutes the framework selected for the numerical analyses presented in this thesis. Coarse grained models considerably reduce the computational time with respect to full atoms simulations, which are not in turn considered in the present setting since typical translocation time scales are yet not accessible, unless the pulling force is order of magnitudes greater than the biological ones [60–62]. In addition, the extrapolation of these results to the lower-force regime is a very difficult task. Moreover, the most recent and remarkable studies refer primarily to the analysis of folding/unfolding pathways [63–65]. The advantage of a reduced complexity with respect to full-atomistic techniques relies on the possibility of massive sampling of events, thus allowing a statistical mechanical description of translocation in terms of ensemble averages, as it will be shown in chapter 3.

Finally, the last section summarizes two numerical investigations of protein transport across nanopores that can be considered as the reference works in the minimalist molecular dynamics setting for protein translocation available in the literature at the beginning of the research activity.

### 2. Experimental work: a selected example from the literature

Stochastic biosensing has been exploited for detection of not only small molecules or DNA strands but also to study long and complex biopolymers. Since our interest is focused on protein experiments and modeling, all the wide literature on oligonucleotides, although extremely interesting, is not further considered.



The transport of a folded polypeptide is characterized by a greater energetic barrier than for an unfolded structure. According with this argument, a recent study combined single-molecule electrical measures with simulations based on Langevin dynamics to show that highly unfolded  $\beta$ -hairpin polypeptides enter the pore in a linear fashion, pursuing fast single-file translocations. In contrast the passage of structured  $\beta$ -hairpin polypeptides occurs more slowly, producing longer current blockades [35].

Similarly, Oukhaled and co-workers [23] studied the translocation of folded, partially folded and unfolded structures of the Maltose Binding protein<sup>1</sup> (MBP) across the  $\alpha$ HL channel (Fig. 2.1). The unfolded structures were accomplished in high concentrations of guanidinium hydrochloride, a chemical denaturant commonly exploited in protein folding analyses. It is worth to disclose that this experimental work constitutes the starting point for the numerical investigations presented here. In this experiment, translocation has been studied by varying the denaturant concentration in the buffer solution, highlighting that current blockades were dependent on this parameter, both in terms of frequency and duration. The authors have observed, among other things, that folded proteins did not translocate across the biopore. The radius of gyration of the native MBP is indeed greater than the  $\alpha$ HL lumen and the driving force was not strong enough to accomplish the protein unfolding at the pore mouth (the voltage required to unzip the fully native MBP would have destroyed the lipid bilayer of the experimental apparatus). Therefore no ion current drops were detected for low levels of denaturant in solution, panels *a* and *c* of Fig 2.1. With the increasing of the guanidinium hydrochloride concentration, the frequency of translocation events increases, following a step-like denaturation curve. Also, the average duration of long blockade events reduces, up to reach the phenomenology depicted in panel *f*, where practically all the MBP molecules are unfolded in the bulk and accomplish translocation in a very short time  $\simeq 1$ ms. In between these states (fully native and fully denatured), complex phenomenologies occur: short and long current blockades coexist, testifying the presence of partially folded structures in solution, the characteristic transport time of which is order of magnitudes greater than the one of unfolded polymers, panels *d* and *e*. The mechanical nature of such bottlenecks of translocation has been deduced since stalls shortened in time when the applied voltage was increased. Moreover, the average translocation time decreases with the increasing of the denaturant concentration. The latter evidence excludes the possibility that the MBP stuck along the pore, due to chemical bonds, a phenomenon that could in principle generate long blockades but that is also thought to increase if the denaturation degree of the molecules

<sup>1</sup>In nature, MBP is responsible for the uptake and efficient catabolism of maltodextrins in the well-known bacterium *Escherichia coli*, increasing the solubility of recombinant proteins when expressed as MBP-fusion proteins. It is a monomeric periplasmic globular protein with two lobes (the *C* and *N* lobes) that close to form the binding site. It is a single chain of 370 amino acids with no disulphide bonds. Its molecular weight is approximately 42.5 ku.

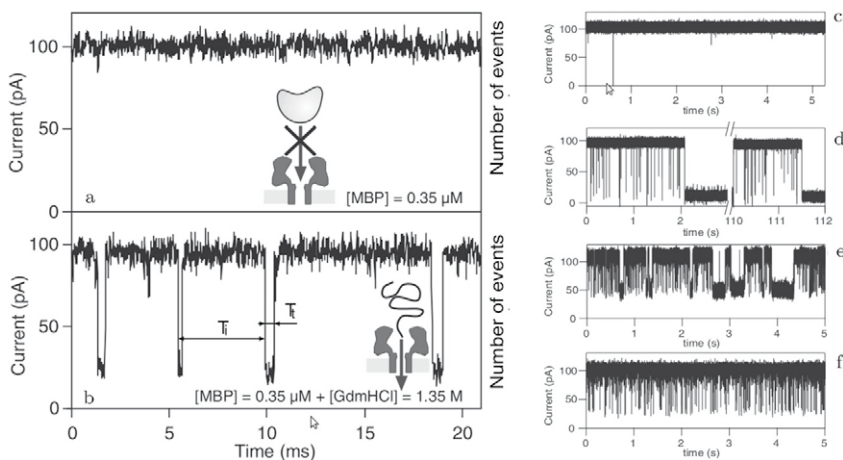


Figure 2.1: Adapted from [23]. Left side: Snapshots of the system and measurements. Panel a: Schematic representation of Maltose Binding protein and  $\alpha$ -hemolysin without Gdm-HCl in solution. The current is constant at  $100\text{pA}$  (open pore conductance) and the MBP can not enter into the pore. Panel b: With  $[\text{Gdm} - \text{HCl}] = 1.35\text{M}$ , well above the unfolding transition,  $[0.8\text{M} : 1.0\text{M}]$ , the current trace decreases down to  $20\text{pA}$  when an MBP molecule is in the channel. Right side: Part of current traces generated by the passage of MBP through the  $\alpha$ -hemolysin pore as a function of the Gdm-HCl concentration. Panel c: At  $[\text{Gdm} - \text{HCl}] = [0.8\text{M}]$  MBP practically does not translocate across the channel and no current drops are detected. Only few short blockades are observed, their average duration is  $0.2\text{ms}$ . Panel d: For  $[\text{Gdm} - \text{HCl}] = 0.85\text{M}$ , long blockades of  $100\text{s}$  in average alternate with a series of short ones of  $0.2\text{ms}$  in average. Panel e: At  $[\text{Gdm} - \text{HCl}] = 0.9\text{M}$ , the frequency of the short blockades increases (of average duration  $0.2\text{ms}$ ). The frequency of the long blockades also increases, their distribution is wide, their mean duration decreases, its value based on the statistical analysis of all events is  $40\text{ms}$ . Panel f: For  $[\text{Gdm} - \text{HCl}] = 1\text{M}$ , only short blockades (of average duration  $0.5\text{ms}$ ) are observed, MBP molecules are almost completely unfolded. The frequency of short events does not further increase.

increases, for the hydrophobic chains of the residues can more easily interact with the pore walls.

As already stated, translocation of Maltose Binding protein through the  $\alpha$ -hemolysin pore is the baseline for the research activity performed. Numerical investigations have the primary aim to explore the complex phenomenology just described in order to point out the role of denaturation on the transport across nanochannels.

### 3. Additional past work outline

The aim of the current section is to touch on additional numerical and experimental works that played a remarkable role during the research activity.

From an experimental perspective, voltage-driven translocation of polypeptides is more problematic than for DNA or RNA strands so that numerical and theoretical support is extremely important. The main difficulties arise from the fact that proteins, differently from nucleic acids, are not uniformly charged molecules and, in addition, their natural tendency to form compact conformations or aggregates usually interferes with the passage through narrow paths, generating consequently a rich (almost overwhelming) collection of non-trivial phenomena [24, 28, 38, 51, 54, 56, 66] (Fig. 2.2). Despite the difficulties, several research groups have showed that voltage-driven

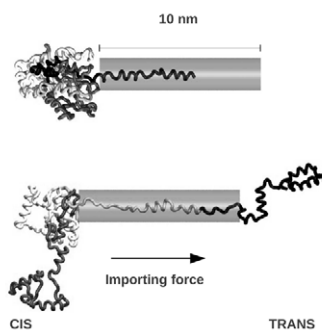


Figure 2.2: Conformations of Maltose Binding Protein (MBP) pulled across a  $\alpha$ -hemolysin channel ( $\alpha$ HL) computed via a reduced-model simulation. The pore is portrayed as a cylinder. The protein description is reduced to its  $C_{\alpha}$ -backbone. The chain needs to unfold to translocate from the cis to the trans-side. The image was made using VMD software [14].

translocation can be achieved, not only for nucleic acids, but also for different peptides and proteins [23, 34, 36, 38, 52, 67–69], and other studies show that this technique is able to yield substantial structural information [70], to characterize protein charge [71] and to discriminate different degrees of folding or aggregation [23]. However, except for some encouraging evidences that nanopores can detect site mutations on proteins [70], it is uncertain whether current blockages convey the necessary information to resolve each of the 20 amino acids in order to achieve protein sequencing.

However, the number of experimental efforts is quickly multiplying [24, 72] due to the obvious biological and biomedical advances expected in this research field [17, 18, 73]. Therefore from both a theoretical [74–79] and computational [11, 80–83] perspective, nanopore technology represents a challenging field to

develop methods able to interpret experimental results and to support the optimization and design of nanopore devices [74–79].

Sometimes simulations and theoretical methods allow bypassing the experimental limitations, such as restricted access to the finer details of the dynamics and irreproducibility of specific phenomena due to the instability of the organic matter out of physiological conditions. There are various theoretical strategies to tackle the transport of proteins across nanopores, starting from the microscopic description. Atomic-scale simulations are extremely informative about the protein dynamics for they take into account the more thorough structural and interaction details of the whole system: protein, nanopore, lipid bilayer and ionic solution [84, 85]. The high system complexity, however, prevents atomic-scale simulations from cumulating the necessary number of events for a meaningful statistical mechanical description, even for small translocating molecules. Simulations via coarse grained models of protein transport across nanochannels constitute an interesting alternative to full-atom methods. As already mentioned, translocation is practically always coupled with unfolding of the biomolecule. The simplest model to account for such a coupled process is to assume that both translocation and unfolding are accomplished by a constant mechanical force. Even this approach is computationally challenging and not devoid of drawbacks, because biologically or experimentally relevant time-scales are orders of magnitude greater than those reproducible by full atoms simulations and, in addition, the constant force assumption could not be realistic to describe the translocation in cells. In fact, typical simulations time scales range from nanoseconds to microseconds (whereas experimentally measured translocation events are of the order of milliseconds). Also, it has been found that repetitive pulling actually catalyzes protein import into mitochondria, suggesting that this could even be a strategy selected by evolution to import proteins into organelles [86]. Specifically, as suggested in [86], the interplay among cyclic operation of molecular motors, force-induced backbone strain, and non-native interactions might be of essential importance in mitochondrial import and not only, as pulling/pushing of proteins by molecular motors is widely observed in the delivery of biomolecules across several organelle membranes.

A simplified phenomenological description of nanopores, proteins and interactions, allocating a reasonable amount of computational resources, provides an immediate inference on the average pathway of the process. Moreover, the limited amount of CPU loading allows a massive statistics of translocation events to be collected. Clearly, the major limitation of coarse grained models lies just in their phenomenological nature, because the non-trivial simplifications introduced have to be compensated by an extensive input of experimental information in order to improve the predictive power of the models and their ability to match experiments. Likely, a suitable strategy to investigate theoretically a complex and articulated phenomenology, as the one emerging from the fast-developing nanopore technology, would require

an integrated approach that cleverly combines coarse-grained modeling with full-atom simulations. In this direction the most famous example of a coarse grained molecular model that directly derives from full-atom force fields (and therefore is much more complex than the approach here actually used) is the Martini model [87, 88].

A burgeoning development of coarse-grained approaches to heteropolymer modeling has taken place in recent years [89]. The basic feature of protein minimalist models is to group together and identify side-chain atoms, or even a complete residue, into simpler units (virtual atoms, beads). These models are simple, fast to implement and require relatively small computational resources as they take into account only the  $C_\alpha$ -carbon backbone dynamics. In this context, the Sorenson Head-Gordon (SHG) model [90] is an off-lattice model that generalizes a previous model introduced by Honeycutt and Thirumalai [91]. The  $\alpha$ -carbons are represented by three possible types of beads: hydrophobic (B), hydrophilic (L), and neutral (N). The force responsible for the collapse onto a compact structure is the attraction between B-beads, whereas all other pairs of interactions are repulsive and determine the rearrangement of the folded structure onto the native topology. The long-range interactions between far apart residues in the sequence are modeled through the potential:

$$V_{LR} = \sum_{i,j \geq i+3}^m \epsilon S_1 \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - 2S_2 \left( \frac{\sigma}{r_{ij}} \right)^6 \right] \quad (2.1)$$

where the summation ranges from 1 to the number  $m$  of residues,  $\epsilon$  sets the energy scale and  $\sigma$  is the excluded volume size used to obtain a self-avoiding polymer chain. Also,  $r_{ij} = r_i - r_j$  is the distance between residues (beads)  $i$  and  $j$ . The attractive forces between hydrophobic beads is attained by setting  $S_1 = S_2 = 1$  for BB pairs. The others are all repulsive: LL, LB are characterized by  $S_1 = 1/3$  and  $S_2 = -1$  and NB, NL by  $S_1 = 1$  and  $S_2 = 0$ . In order to achieve the proper secondary structure, bending and dihedral interactions, which surrogate side-chain packing and hydrogen-bonding, are introduced. The analytic expression of the dihedral potential is

$$V_{dih} = \sum_{i=1}^{m-3} [A_i(1 + \cos \phi_i) + B_i(1 - \cos \phi_i) + C_i(1 + \cos 3\phi_i) + D_i(1 + \cos(\phi_i + \pi/4))], \quad (2.2)$$

where  $\phi_i$  indicates the angle between the two adjacent planes identified by the positions of four consecutive beads. The information on secondary structures is stored in the coefficients  $A, B, C$ , and  $D$  that determine a bias on the angles reflecting the tendency of residues to form one of the secondary motives: helical (H), extended (E), turn (T). Therefore, the primary structure must be complemented with the auxiliary sequence of E, H or T symbols, assigning the appropriate set of coefficients. Such a decoupling between primary and dihedral sequence

allows a fine structural design [92]. The bending angle interaction is modeled as a harmonic potential:

$$V_{bend} = \sum_{i=1}^{m-2} \frac{k_{\theta}}{2} (\theta_i - \theta_0)^2, \quad (2.3)$$

with a constant  $k_{\theta} = 20\epsilon^2/(rad)$  so that large deviations from the equilibrium value  $\theta_0 = 104.85^\circ$  are unlikely. The SHG force field is completed by stiff springs with equilibrium distance  $d_0 = 3.8 \text{ \AA}$ , namely

$$V_{pept}(r_{i,i+1}) = \sum_{i=1}^{m-1} \frac{k}{2} (r_{i,i+1} - d_0)^2, \quad (2.4)$$

used to maintain chain connectivity by simulating the presence of covalent peptide bonds between successive amino acids. In the SHG model a selected number of elements is used to capture the essential characteristics of the proteins. However, some features, such as hydrogen bonding and side chains, are missing. These limitations should be compensated through a design strategy for optimizing the sequence [93]. SHG models have been adopted variously in studies on protein unfolding (in free and nano-confined geometries [54]) and translocation [66]. In the next chapter the coarse grained model (Gō model) used to perform the research activity here reported is considered, along with all the other assumptions employed. A critical comparison between the performance of the Gō and SHG models in reproducing the folding of the small WW-domain form in the HPin1 protein can be found in [94].

Huang, Kirmizialtin and Makarov [54], using a SHG force field, compare the mechanical unfolding of Ubiquitin in free space (AFM<sup>2</sup>-like stretching) and in a semi-infinite pore, and show that the two dynamics are quite different. Moreover, the authors observe that the unfolding pathway in the pore is strongly dependent on both pore diameter and pulling terminus. The free energy profile as a function of the position of the pulled amino acid along the pore axis presents several branches corresponding to ‘unfolding intermediates’ before a plateau is reached once the whole protein has entered the pore. As we shall see, similar intermediates are also observed as ‘stall events’ in the present research activity in a finite-size pore [56]. Due to its structural simplicity, Ubiquitin was used also in other studies, for instance Ammenti et al. [55] simulated Ubiquitin translocation across a long pore using the Gō force field. The latter two examples [54, 55] are further considered in the next section as are deemed to constitute the reference coarse grained computational studies available at the beginning of the research activity. A brief introduction is provided to delineate the basic mathematical

<sup>2</sup>Atomic Force Microscopy

frameworks that are still missing to perform an albeit qualitative description of these works.

#### 4. Analytical models and simulations of protein translocation

##### 4.1. Overview

A common analytical approach to model complex phenomena such as the translocation process, is to investigate the low-dimensional potential of the mean force landscape, which is relevant to suitably-defined reaction coordinates. This approach is often coupled with simulations where simplified models for the protein and its surroundings are exploited, both to investigate the dynamics of the system and to reconstruct the free energy profile, which governs the trend of the potential of the mean force as it will be clear later on. As developed further in chapter 4, a reaction coordinate is just a collective order parameter that encompasses and summarizes multi-dimensional phenomena, usually between two end-states. The choice of the reaction coordinate is often guided just by intuition and aim of the analysis. In forced dynamics a natural selection is to choose the degree of freedom that couples with the driving force (usually it is the component along the force direction of the position vector of the pulled residue of the protein). For *umbrella sampling* applications [95] to translocation, however, the center of mass of the molecule is often the preferred parameter [55]. For mechanical stretching, a situation mimicking AFM pulling studies, the protein extension along the direction of the applied force is often implemented as a reaction coordinate. When the free energy landscape enters in the formulation of the potential of the mean force, which is the natural (and sole) option for canonical systems such as a single molecule in a heat reservoir, an implicit assumption is always done: the translocation can be presumed to take place in thermal equilibrium at any location along the pore, so to obtain a quasi-equilibrium process. It is however possible that a protein, while inside the tunnel, would be trapped in a metastable state so that its behavior is practically non-ergodic. The quasi-equilibrium assumption allows one to avoid dealing with the actual kinetics of the translocation, that can be traced back to a series of quasi-static states, as the ones obtained by sampling the system with the umbrella sampling method. In this context the free energy trend,  $G_0(z)$ , (here the reaction coordinate is assumed to be just one for simplicity and it is indicated by  $z$  without loss of generality) is related to the probability  $p(z)$  to find the chosen reaction coordinate at  $z$  (a collection of ‘equivalent’ macro-states of the system correspond to such a value of the collective variable) by the following relation

$$G_0(z) = -k_b T \ln p(z), \quad (2.5)$$

where  $k_b$  is the Boltzmann constant and  $T$  the temperature of the heat reservoir. When external fields act on the system, the associated reversible work  $W(z)$  has to be taken into account by including in the previous expression a proper term. For instance, in the case of the mechanical stretching induced by a constant force  $f$ , exemplified in the next paragraph, the potential of the mean force reads as

$$G_f(z) = G_0(z) + W(z) = -k_b T \ln p(z) - fz \quad (2.6)$$

(and now it is evident the reason why the free energy profile governs the potential of the mean force trend). Since continuum approaches are delved deeper later on in the thesis, the topic is not further considered in the present section. It has had the only aim to introduce the idea underneath the 1D continuum approach related to the already mentioned numerical investigations [54, 55]. Such a description is carried out in the next section.

#### 4.2. Coarse-grained molecular dynamics reference studies

Huang, Kirmizialtin and Makarov [54] performed an analysis relevant to protein import into mitochondria using Langevin dynamics simulations (see (3.8b)) of a coarse-grained off-lattice model, the already introduced SGH model, to investigate the co-translocation unfolding of a protein structure (the 76 amino acids long Ubiquitin), pulled mechanically through a semi-infinite narrow pore, cylindrical in shape. It is emphasized that, despite the computational savings, due to the minimalist assumptions, biologically relevant time scales associated with barrier-crossing events are rarely accessible via direct simulations. In this work the potential of the mean force just introduced is exploited, *i.e.* equation (2.6), and the chosen reaction coordinate coincides with the  $z$ -axis coordinate of the pulled residue (the  $z$ -axis coincides in turn with the axis of symmetry of the pore). Thus the translocation is assumed to be slow enough with respect to the protein internal dynamics to be described as a one-dimensional driven diffusion along the pore axis in the potential  $G_f(z)$ . To obtain the global shape of the free energy  $G_0(z)$ , the umbrella sampling/weighted histogram method [96] is used.

In [54] the free energy profile resulting from translocation is also compared with the one stemming from mechanical stretching (AFM-like experiments) simulations and differences are discussed. It is in fact proven that significant differences exist between these two cases and that, in particular, AFM experiment outcomes cannot be easily scaled to match translocation results as each process occurs following distinct pathways. For details on amino acid interactions see section 3. Without entering the details of the calculations, Figure 2.3 panel *a* shows the profile of the potential of the mean force in the mechanical stretching case, for different values of the force (here the reaction coordinate is the molecule extension along the force direction). In this figure, contact maps are also depicted. These are the plots of the pairs



of residues  $(i, j)$  such that their distance, in both the native and current configurations, is lower than (in this analysis)  $7.5\text{\AA}$  and  $9.125\text{\AA}$  respectively. Also,  $|j - i| \geq 3$ . Such residues interact following the long range potential (2.1). In panel *b* the same situation is showed for the translocation process when the *N*-terminus is pulled inside the pore. In the article [54] the pulling from the *C*-terminus is also studied<sup>3</sup>. While the overall free energy cost of extending the

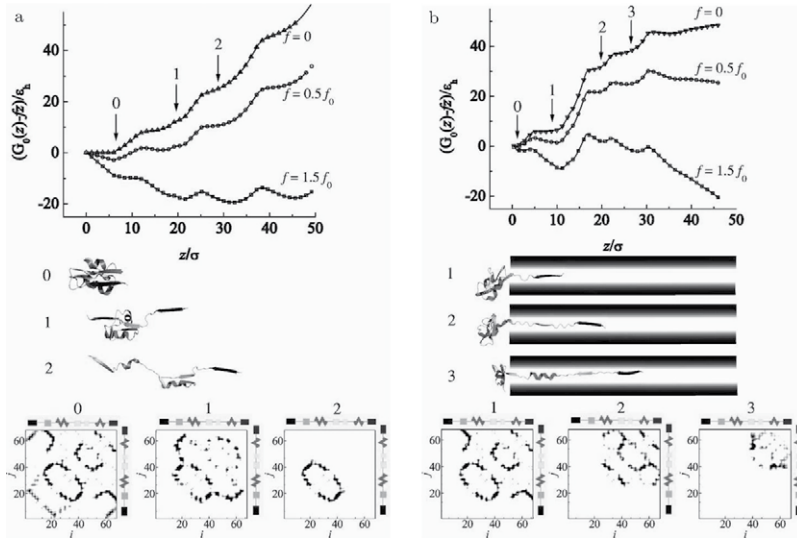


Figure 2.3: Adapted from [54]. Panel a: Free energy profile of Ubiquitin in units of  $\epsilon_h$  as a function of the mechanical stretching reaction coordinate for different values of the force. The minima correspond to the native-like state 0 and unfolding intermediates 1 and 2, which structure is depicted in the lower part of the figure, where also contact maps are showed. Panel b: Translocation of Ubiquitin driven by a force applied to its *N*-terminus. The potential of the mean force is plotted as a function of the position of the reaction coordinate, corresponding to the just mentioned end of the molecule. Native-like 0 and intermediate structures 1-3 are shown with the relevant contact maps.

protein is similar, the profiles of  $G_f(z)$  and, consequently, the force-induced unfolding mechanisms are different. The local minima correspond to unfolding intermediates that are obviously distinct. Certain similarities exist between mechanical stretching and translocation pathways, especially in the first step,

<sup>3</sup>The *N*-terminus refers to the start of a protein or polypeptide terminated by an amino acid with a free amine group ( $-\text{NH}_2$ ). The convention for writing peptide sequences is to put the *N*-terminus on the left and write the sequence from *N*- to *C*-terminus. When the protein is translated from messenger RNA, it is created from *N*-terminus to *C*-terminus. The latter is in turn the end of an amino acid chain (protein or polypeptide), terminated by a free carboxyl group ( $-\text{COOH}$ )

transition from structure 0 (native-like) to structure 1, which is present in both cases and concerns the same event (separation of two terminal parallel strands).

In this study the dependence of the translocation time on the driving force is also examined, in addition to pore-size effects. Actually only a crude estimate of the effect of the driving force on the blockage time is carried out, as the authors assume that the average translocation velocity should be correlated to the overall free energy barrier, namely  $\Delta G(f) = \max G_f(z) - \min G_f(z)$ , encountered along the reaction coordinate. The unfolding free energy barrier does not depend linearly on the force, as it is often conjectured in mechanical stretching studies. The pore investigation is not considered here for brevity; it is just worth mentioning that if the pore is wide enough, the protein can perform transloca-

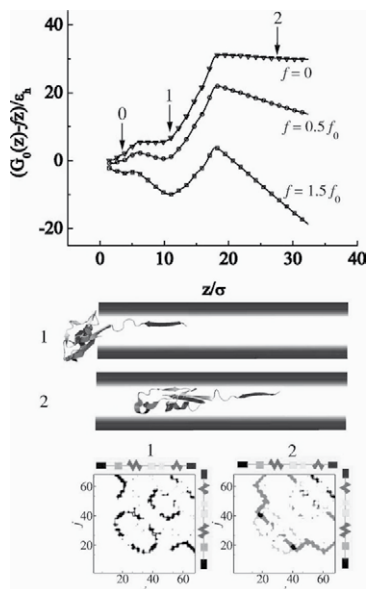


Figure 2.4: Adapted from [54]. Translocation of Ubiquitin across a large pore. The reaction coordinate used to project the free energy profile is the position of the last amino acid of the chain along the pore, the  $N$ -terminus. Structures representative of the molecule as encountered during simulations are also depicted with their contact maps.

tion in a semi-folded fashion, for example in the way showed in Figure 2.4.

As stated by the authors themselves, several limitations may prevent the direct comparison of the numerical results with the experimental ones. These are subsequently listed:

- The minimalist protein model may be too rough to reliably estimate the translocation free energy barriers and their force dependence.
- The analysis concerns a semi-infinite pore, which is unrealistic.

- The constant force assumption might be inadequate, even for voltage-driven translocation experiments.
- The assumption that translocation is a slow barrier crossing process may be unrealistic, especially in the case where the pore is wide and the resulting free energy barrier is small. If the translocation time scale is of the same magnitude of internal dynamics time scales, the simple description of the one-dimensional driven diffusion process in the equilibrium potential of the mean force is no longer applicable and a full simulation of translocation dynamics may be required.
- More realistic pore models may be called upon in order to apply computational results to the natural phenomena that take place in cells.

For completeness, it is worth mentioning a work where translocation of Ubiquitin is studied both in its thermodynamics and average kinetics [55]. In this analysis a pore of finite length is considered, thus tackling the issue, although it is long enough to accommodate the unfolded protein, *i.e.* if some residues still dwell on the cis-side, none has yet escaped the channel from the trans-side aperture. In [55] statistical analyses are also carried out, accounting for blockade time distribution and translocation probability, by following the analytical method just sketched, in terms of first passage time statistics (again this perspective will be delved deeper further on in the thesis, chapter 4 sec. 2.1). In the remaining part of this section the main points of [55] are summarized. The model employed for the protein is a Gō-like one, see sec. 2.1 of chapter 3, while the pore confinement is again described by a cylindrical soft-core repulsive potential. The Langevin equation of motion still governs the simulation dynamics and umbrella sampling is performed by constraining the center of mass of the molecule with a harmonic potential. Multiple weighed histogram technique is used to deweight the umbrella sampling results [96].

Translocation simulations are run until the protein is fully expelled out of the right end of the channel and almost complete refolding takes place. However, at low forces, Ubiquitin may fail to cross the pore within the simulation time window. The probability of translocation,  $P_{T_r}$ , is estimated as the number of translocation success over the total number of runs as a function of the driving force and of temperature. Figure 2.5, panel *a*, shows the curves of  $P_{T_r}$  for two different values of the temperature, reporting a step-like trend, which delineates a clear value of the critical force, *i.e.* the force for which  $P_{T_r} = 1/2$ . For instance, the decreasing of the critical force with temperature is a tangible consequence of the lowering of the free energy translocation barriers with temperature increase. Panel *b* in the same figure contains the trend of the average translocation time as a function of the pulling regime and for the two values of the temperature simulated. Only in the high force regime the average translocation time shows an Arrhenius-like dependence on the force, *i.e.*  $\tau \simeq \tau_0 \exp(-\beta FL)$ , with  $\beta = 1/(k_b T)$ ,  $F$  the pulling force and  $L$  the pore length.

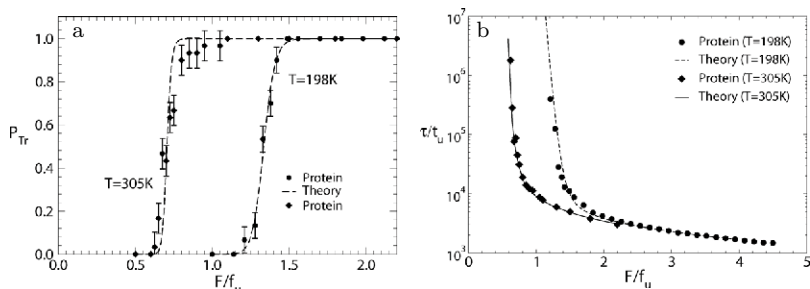


Figure 2.5: Adapted from [55]. Panel a: Ubiquitin translocation probability as a function of the pulling force for two values of temperature. Pore length = 300 Å, pore radius = 4 Å,  $f_u = 6$  pN,  $T_{ref} = 198^\circ\text{K}$  and  $T_{ph} = 305^\circ\text{K}$ . Panel b: Average translocation time  $\tau$  of Ubiquitin as a function of the force  $F$  at the two simulated temperatures.  $t_u = 0.25$  ps.

Characteristic trends of the free energy (indicated as  $G(x)$ ), obtained from umbrella sampling simulations and multiple weighted histogram analysis method, are reported in Figure 2.6. These values refer to the position of the center of mass of the protein along the pore axis. The plateau indicates that once the protein is unfolded and completely inside the channel, it can slide along its axis without further free energy increase. This is a consequence of the pore geometry, for it is long enough to fully host the protein even when completely unfolded. In contrast, as expected, the main variations of the free energy occur at pore boundaries. In the proximity of the right end, when actually some of the residues are already located outside the pore, the molecule starts to be spontaneously expelled.

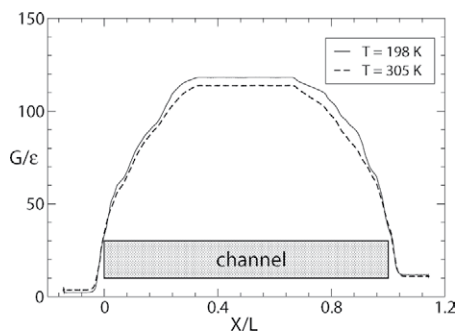


Figure 2.6: Adapted from [55]. Free energy trend obtained from umbrella sampling simulations with a harmonic potential as umbrella potential in units of  $\epsilon$ . The two curves refer to the two simulated temperatures (reference and physiological).

## Coarse-grained molecular dynamics and continuum models

However, despite the insights gained, the general relationship between the structure of the proteins and their resistance to mechanically driven co-translocation unfolding remains poorly understood.

## Chapter 3

# Molecular Dynamics Translocation Simulations

### 1. Preamble

A set of translocation simulations and the description of the relevant numerical tools are reported here. The attention is focused on driven transport of Maltose Binding protein (MBP) across nano-channels in the framework of coarse-grained modeling.

As already mentioned, this work is motivated by a recent experiment on the MBP voltage-driven transport across nanopores which explores the influence of denaturation on translocation pathways [23]. Specifically, short and long channel blockades, associated with the translocation of denaturated and partially folded MBP conformations respectively, are identified by large differences in the ion current drop durations.

Dynamically, translocation of proteins strongly depends on the denaturation degree and generally is coupled with an unfolding stage. In fact a folded protein with gyration radius larger than the pore narrower section needs a complete or partial unfolding to start the translocation [38, 51].

Structurally, MBP is a monomeric globular protein with 370 residues, resolved through x-ray crystallography by Quioco, Spurlino and Rodseth [97]. Mechanical AFM pulling experiments by Bertz and Rief [98] identified some structural regions - named *unfolds* by the authors - as resistant areas to mechanical denaturation. The domains M1, M2, M3 and M4, shown in Figure 3.1A on the PDB structure (PDB-ID: 4mbp), can be related (one can expect such a thing) to the long blockade events in the MBP translocation. In computations we consider for MBP the Gō-like model described in section 2.1, as a natural approach to assess the impact of the molecule structural properties on translocation. The advantage of a coarse-grained description with respect to atomic scale models relies on the possibility to explore a large number of denaturation and pulling conditions, accumulating this way robust statistics of translocation events. For recent applications of coarse-grained models to various biochemical processes see the review [99].

As first step we show that the model is able to reproduce the general features of MBP chemical denaturation. Afterward we study the translocation process in a pore reproducing the average dimensions of the  $\alpha$ HL channel. We find the translocation dynamics to be strongly affected by the protein denaturation state,

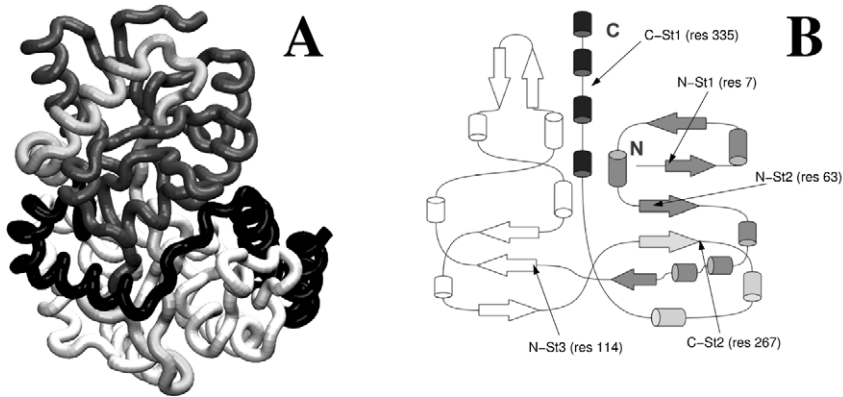


Figure 3.1: Adapted from [98]. A) Maltose Binding Protein structure with the four unfoldons highlighted. M1-darkest gray (residues: 296-370), M2-light gray (residues: 244-295), M3-dark gray (residues: 1-113), M4-lighter gray (residues: 114-243). Panel B reports the positions of the five stall-points found in our translocation runs (residues 7, 63, 114, 267, 335) in a 2D-topological view.

similarly to the experimental evidence. In particular, the translocation of chemically unfolded MBP conformations requires relatively low forces and once the pulled terminus enters the pore, the transport proceeds uniformly. In contrast, native-like structures exhibit a richer phenomenology: stronger forces are required to trigger the transport that, once started, develops in a *stick-and-slip* fashion, through bottlenecks and jerky movements caused by the rearrangements of the folded part of the protein that has not yet engaged the pore. In this case the issue is to identify the MBP structural motives responsible for the translocation slowing down. First we exclude these stalling stages to be related with the unfoldons, by showing that translocation pathways significantly differ from the free space unfolding dynamics. Then, by means of an analysis of static and dynamic native contact maps, we show that the stall points of the translocation pathway are mainly due to the protein regions more dense in long range native interactions. This result might constitute a possible criterion to predict *a-priori* some statistical features of protein translocation from a simple static structural analysis.

## 2. Three-dimensional models for MBP simulations

In order to simulate numerically translocation phenomena, a coarse-grained computational model for the Maltose Binding protein and its surroundings is developed (Fig. 3.2). As it is described below, choices are made in terms of the importing mechanism, pore and particle interaction potentials, numerical integration scheme and equation of motion for the material points constituting the

system. Finally, the approaches followed in the practical implementation of the numerical simulations are also expounded.

### 2.1. MBP numerical model: The Gō-like approach

The phenomenological off-grid model proposed by Nobuhiro-Gō and Harold A. Scheraga (minimalist off-lattice native-centric  $C_\alpha$  Gō-model) [1, 100] is a coarse-grained model that identifies all the amino acids with their  $C_\alpha$  carbon atoms (Fig. 3.2 panel *c*). Therefore, the protein is reduced to a sequence

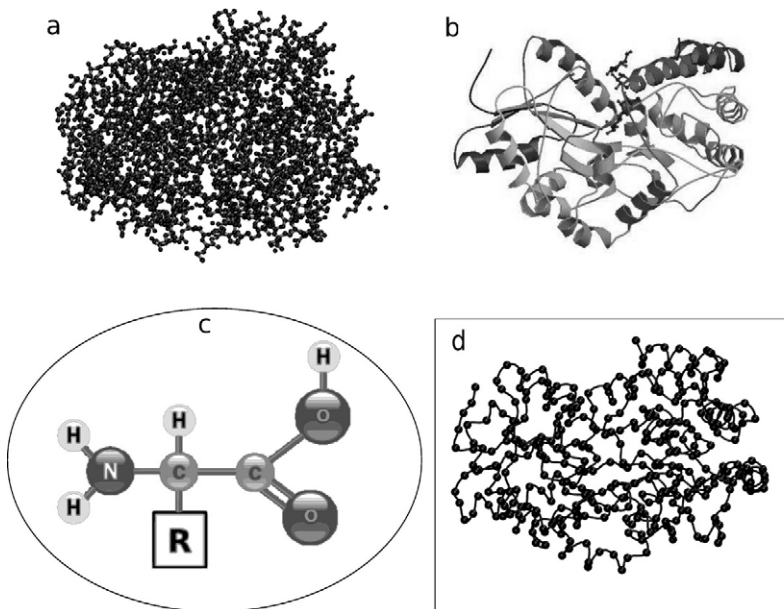


Figure 3.2: Panel a: Maltose Binding protein appearance according to a full atom representation. Panel b: Secondary structure of the MBP evinced from its protein data bank file, PDB-ID: 4mbp. Panel c: Amino acid core structure. The characteristic residue chain of this elementary constituent is summarized by a box surrounding a capital R. The  $C_\alpha$  carbon atom binds directly with the residue chain. Panel d: MBP coarse grained representation. Accordingly to the Gō model, only the  $C_\alpha$  carbon atoms along the main backbone chain of the protein are depicted.

of material points (also called beads) coinciding spatially with the  $C_\alpha$  atoms of the protein backbone and no side chain characteristics are retained (panel *d* of the same figure). The atomic mass is the same for all the residues and equal to 1 in the so-called ‘code measure units’ or ‘internal units’.



A lot of variances and refinements of this model are available in the current literature (see e.g. [2, 101–105]). All these approaches are generally referred as Gō-like models. Here the approach presented in [2] is implemented.

The minimum energy is assigned to a particular reference configuration, usually coincident with the native structure of the considered protein. This is why the Gō model can be properly applied only when the crystallized structure is already known and the folding dynamics are mainly determined by the topology of the nated conformation, as it happens with small globular proteins. The Gō-like model generates an almost ideal folding pathway (energetic traps associated with metastable conformations able to hamper the folding process are practically absent). Also, the native configurations so obtained are stable and close to the reference one (in terms of root mean square deviation [106]), which can be extracted, for example, from the protein data bank. It is a particularly suitable approach to simulate two-state folding dynamics, but it has been used also to predict intermediate configurations [2].

Thanks to its properties and to its simplicity the Gō model has been widely exploited for protein numerical analysis [55, 94, 107–115]. The main limitation of the approach is the requirement that the native tertiary structure of the protein has to be known. In addition, its topological nature does not account for chemical factors, determining a considerable disagreement between real and simulated time scales: kinetic peculiarities of the folding process are not properly predicted. These discrepancies do not alter the folding thermodynamics and the correct sequence of events. In most cases the model allows the correct assessment of the free energy values and of folding pathways. Moreover, transition states are coherent with the ones experimentally determined.

In the remaining part of this section the practical implementation of the model is presented. Following the Gō-like approach selected, the potential acting on the residues of the MBP is constituted by four parts.

- Peptide potential (or bond potential),  $V_p$ .
- Bending angle potential,  $V_\theta$ .
- Twist angle potential,  $V_\phi$ .
- Non-bonded interactions (Lennard-Jones potentials or barriers),  $V_{nb}$ .

Let  $r_i, r_i^0$  (with  $i = 1, \dots, m$ ) be the position vectors of the  $m$  residues identified by their  $C_\alpha$  carbon atoms, in the reference (native) and current configurations, respectively. The peptide potential, responsible for the covalent bonds between the beads of the polymer chain, has the following representation:

$$V_p(r_{i,i+1}) = \frac{k_p}{2} (r_{i,i+1} - r_{i,i+1}^0)^2, \quad (3.1)$$

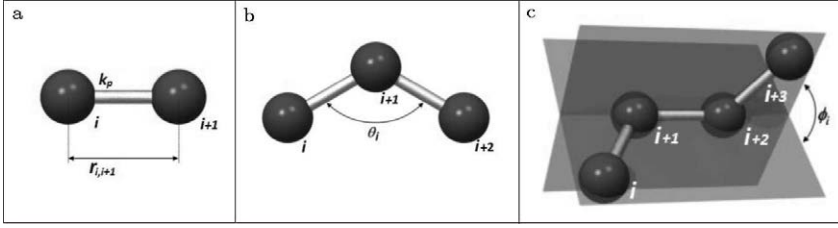


Figure 3.3: Panel a: Schematic view of the peptide bond. Panel b: Sketch of the bending angle interaction. Panel c: Representation of the twist angle formed by the two planes determined by four consecutive material points of the protein structure (three by three).

where  $r_{i,i+1} = r_i - r_{i+1}$ ,  $r_{i,i+1}^0 = r_i^0 - r_{i+1}^0$  are the position vectors (their norm is the bond length) between the bonded residues  $i$  and  $i+1$  in the instantaneous and native configurations, respectively (panel a of Fig. 3.3). Moreover,  $k_p = 1000\epsilon/d_0^2$  is the empirical constant in terms of the parameters  $d_0$  and  $\epsilon$  named, respectively, the *equilibrium length* and the *unit of energy*. In particular,  $d_0 = 3.8\text{\AA}$  is the average distance between two adjacent amino acids and  $\epsilon$  sets the energy scale of the model. In summary, the bond potential is nothing more than a simple harmonic potential.

The angular bending potential  $V_\theta$  plays a fundamental role in recovering the secondary structure of the reference native configuration of the protein. It is mathematically equivalent to the peptide potential, where relative displacements are substituted by angular differences, namely

$$V_\theta(\theta_i) = \frac{1}{2}k_\theta(\theta_i - \theta_i^0)^2, \quad (3.2)$$

where  $k_\theta = 20\epsilon \text{ rad}^{-2}$  is the elastic constant expressed in terms of the energy computational unit and  $\theta_i$ ,  $\theta_i^0$  are the bond angles formed by three adjacent beads in the simulated (time-dependent) and native conformations, respectively (Fig. 3.3b).

An additional contribution to the composition of the secondary structure is the dihedral potential, expressed as a function of the twist angles  $\phi_i$  and  $\phi_i^0$ , again referred respectively to the actual and crystallized conformations. The twist angle is the angle formed between the two planes determined by four consecutive amino acids (three by three) along the chain, panel c. The definition of the twist angle potential  $V_\phi$  is

$$V_\phi(\phi_i) = k_\phi^{(1)}[1 - \cos(\phi_i - \phi_i^0)] + k_\phi^{(3)}[1 - \cos 3(\phi_i - \phi_i^0)], \quad (3.3)$$

where  $k_\phi^{(1)} = \epsilon$  and  $k_\phi^{(3)} = \epsilon/2$  are the dihedral constants.

In the Gō model a distinction is made among the pairs of residues that interact following a potential that has also an attractive part, in addition to a repulsive

one, set to take into account excluded volume effects, and the ones that, instead, only repulse each other sterically. The distinction is made on the basis of the native distances with respect to a parameter of the model, the so-called ‘cut-off radius’,  $R_c$ . If the native distance between two ( $i$  and  $j$ ) non-adjacent residues, *i.e.*  $j > i + 3$ , is lower than the cut-off distance, they are considered to be in native contact and will interact following a 12-10 Lennard-Jones potential (3.4). In other words, such residues will interact attractively if brought to a distance greater than the native one and repulsively otherwise:

$$V_{nat}(r_{ij}) = \epsilon \left[ 5 \left( \frac{r_{ij}^0}{r_{ij}} \right)^{12} - 6 \left( \frac{r_{ij}^0}{r_{ij}} \right)^{10} \right], \quad (3.4)$$

where all the symbols have been already defined. When  $r_{ij}^0 > R_c$ , the purely repulsive contribution  $V_{nnat}$  is assigned to the pair of amino acids considered. This term is based on excluded volume effects. It enhances the cooperativity in the folding process and takes the form of a Lennard-Jones barrier

$$V_{nnat}(r_{ij}) = \frac{10}{3} \epsilon \left( \frac{\sigma}{r_{ij}} \right)^{12}, \quad (3.5)$$

where  $\sigma$  is, similarly to the SHG model, a free length parameter correlated with the extension of the excluded volume (self-avoiding polymer chain). It is usually set equal to the average dimension of the amino acids, approximately 4.5Å.

The total potential acting on all the residues of the protein is then:

$$V_{G\bar{o}}(r_{ij}) = \sum_{i=1}^{m-1} V_p(r_{i,i+1}) + \sum_{i=1}^{m-2} V_\theta(\theta_i) + \sum_{i=1}^{m-3} V_\phi(\phi_i) + \sum_{i,j>i+3} V_{nb}(r_{ij}), \quad (3.6)$$

where the non-bonded term  $V_{nb}$  summarizes the possible long range interactions just described and reads as

$$V_{nb}(r_{ij}) = \epsilon \begin{cases} V_{nat} = \left[ 5 \left( \frac{r_{ij}^0}{r_{ij}} \right)^{12} - 6 \left( \frac{r_{ij}^0}{r_{ij}} \right)^{10} \right] & r_{ij}^0 < R_c, \\ V_{nnat} = \frac{10}{3} \left( \frac{\sigma}{r_{ij}} \right)^{12} & r_{ij}^0 > R_c. \end{cases}$$

In what follows  $\epsilon = 1$  and  $R_c$  varies in the range (3.0 ÷ 7.5)Å.

## 2.2. Pore and pulling models

The confinement effect on a protein driven inside a narrow space such as a nanopore can be represented by a step-like soft-core repulsive cylindrical po-

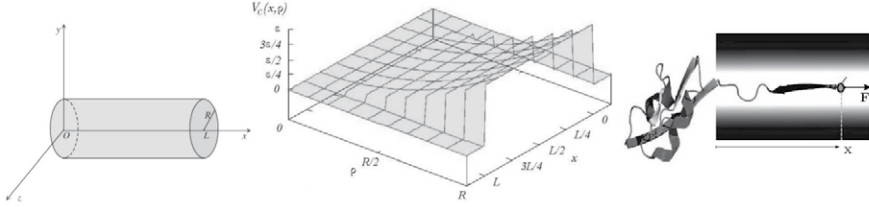


Figure 3.4: A step-like  $[0 : L]$  cylindrical soft-core repulsive potential  $V_c$  is chosen to model the effect of the confinement inside the pore. A constant force is used to drive the MBP inside the channel. It is always applied to the foremost residue inside the pore.

tential. For simplicity the cylinder axis of symmetry is set coincident with the  $x$ -axis of the frame of reference used for translocation simulations.

The same direction is used to develop the mechanical pulling of the MBP by the application of a constant force to the foremost residue inside the channel. The constant action is thought to constitute a reasonable approximation for the average effect of the electrical potential in a voltage-driven translocation experiment.

The channel potential is given in (3.7) and schematic representations of the system, the potential and the pulling mechanism, are shown in Figure 3.4.

$$V_c(x, y, z) = V_0 \left( \frac{\rho}{R_p} \right)^{2q} \Theta[x(L - x)]. \quad (3.7)$$

Here  $\Theta(s) = [1 + \tanh(\alpha s)]/2$  is a smooth step-like function limiting the action of the pore potential in the effective region  $[0, L]$ .  $L$  and  $R_p$  are pore length and radius respectively. Also,  $\rho = \sqrt{(y^2 + z^2)}$  is the radial coordinate. A convenient choice of the other parameters is  $q = 1$ ,  $\alpha = 3\text{\AA}^{-2}$  and  $V_0 = 2\epsilon$  [114]. The driving force acts only in the region in front of the pore mouth (the capture one)  $x \in [-2, 0]$ , and inside the channel  $[0, L]$ . Pore length  $L = 100\text{\AA}$  and radius  $R_p = 10\text{\AA}$  are taken from  $\alpha$ HL structural data.

### 2.3. Langevin dynamics for the material points of the MBP

The choice for the dynamics of the beads is not different from what is usually employed in coarse-grained molecular dynamics simulations for it is the well-known Langevin equation (3.8a) that governs the motion of the material points. Consequently, numerical investigations are performed at constant temperature. The overdamped limit is actually implemented, i.e.  $\dot{r} = 0$ , (3.8b); a standard Verlet algorithm is used as numerical scheme for time integration [116]. In other words, any particle of the system follows a dynamics enforced by this

three-dimensional equation, when a numerical simulation runs:

$$\ddot{r}_i = -\gamma \dot{r}_i - \nabla_{r_i} V(r_i) + \Gamma(t) + F_{ext}, \quad (3.8a)$$

$$\dot{r}_i = \frac{-\nabla_{r_i} V(r_i) + \Gamma(t) + F_{ext}}{\gamma}. \quad (3.8b)$$

Here  $\gamma$  is the friction coefficient used to keep the temperature constant (also referred to as Langevin thermostat) and  $\Gamma(t)$  takes into account thermal fluctuations, being a delta-correlated stationary and standard Gaussian process (white noise). Practically, the latter quantity represents a random force that, along with the friction term, satisfies the fluctuation-dissipation theorem (namely the mean-square value of  $\Gamma$  is proportional to the corresponding friction coefficient,  $\gamma$ ).  $V(r_i)$  is the total potential on the material points (generally speaking it includes both the terms belonging to the Gō-like model and to the potential used to model the pore confinement (3.6) and (3.7), respectively, depending on the position  $r_i$  of the issued bead) and  $F_{ext}$  takes into account the possible contribution of the external forces (in the present setting it coincides with the force driving the translocation, which acts only on the pulled residue).

Several analytical models can be implemented starting from a Langevin approach but, since this topic is developed when 1D-continuum modeling is considered, it is not further discussed here.

### 3. Methods for translocation and stretching simulations

The following sections describe the practical implementations of the numerical investigations in the cases of mechanical pulling in a confined environment (translocation simulations) and in free space (AFM-like mechanical stretching simulations). Thermalization runs are also outlined when necessary as they have been used to obtain statistically independent initial configurations.

#### 3.1. Translocation simulations

For each cut-off radius  $R_c$ , thermalization simulations are performed at  $T = 0.75 \theta_u$  in the already mentioned ‘code measure units’ (symbols will be usually dropped in what follows) to generate initial MBP configurations with one of the two ends constrained near the pore entrance. More in general, in a thermalization (or equilibration) simulation the protein is left free to fluctuate and possibly to unfold (depending on the value of the cut-off radius) in the heat reservoir. Protein configurations are sampled at time intervals equal to 10% of the simulation time window,  $T_w = 10^5 t_u$  with  $t_u$  time units, to ensure statistical independence of initial states. After thermalization, the protein is driven by applying the pulling force  $F$  in the capture region and into the pore.

Protein translocation is considered accomplished when the last residue leaves the trans-side of the channel. As it will be shown in the result section, there are conditions where the molecule escapes the capture by diffusion, despite the action of the importing force. To save computational time, simulations are stopped and discarded if the protein reaches a distance from the pore entrance comparable to its native state linear size (approximately  $20\text{\AA}$ ). This criterion is based on a preliminary estimate of a negligible probability for the molecule to re-approach the pore by diffusion and re-start the translocation.

To set the nomenclature up, such cases of protein escape will be labeled as *loss* events. Cases where the protein is neither translocated nor *lost* at the end of the time window are indicated as *unsuccess*.

### 3.2. Stretching simulations

Before each stretching simulation (AFM-like), the protein is equilibrated without constraints or additional forces in order to obtain a proper initial configuration. Stretching is simulated by using a constant velocity Steered Molecular Dynamics (SMD) strategy [117] where protein elongation is induced by a spring of elastic constant  $k = 0.1\epsilon$ . The protein N-terminus is held fixed, the C-terminus is attached to the first end of the spring, which second end is dragged at constant velocity  $v$  in the direction of the initial end-to-end vector. To test the resulting robustness of simulation protocols, we use different steering velocities and perform AFM-like stretching also from the N-terminus (with the C-terminus blocked).

As mentioned in the introduction, MBP contains four unfoldons (M1, M2, M3 and M4, Fig. 3.1). To monitor the denaturation degree of the  $k$ -th unfoldon domain, we consider the number of its internal active native contacts, normalized to the corresponding value in the PDB crystallized structure, as a function of time. Two residues originally in contact are considered detached when their distance in the actual molecule conformation exceeds  $1.23r_{ij}^0$ . The latter criterion is based on a negligible attractive force beyond such a distance, which approximately coincides with the point where the 12-10 Lennard-Jones potential (3.4) changes curvature.

## 4. Translocation and stretching numerical results

### 4.1. Denaturation characterization

In the Gō-model force field (3.6) a decrement in the cut-off radius  $R_c$  plays virtually the rôle of a chemical denaturing agent because, by reducing the number of attractive long-range interactions (3.7), the native state is destabilized with respect to the action of thermal fluctuations. Denaturation can be also achieved by increasing the temperature or decreasing the energy scale  $\epsilon$  (thermal denat-

uration). The latter approach would be however too *global* to be compared to specific experiments [23, 27], because it would soften both bonded and non-bonded interactions (see eqs. (3.6), (3.7)). The denaturation we implement in

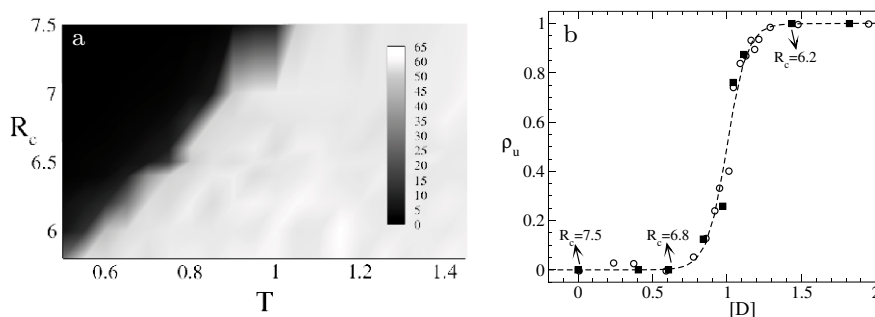


Figure 3.5: Panel a: Contour map of the RMSD as a function of the cut-off radius and simulation temperature. Darker regions reflect states where the equilibrated protein structure resembles the native configuration. The transition from folded to unfolded state is a two-state transition and is extremely sharp, in compliance with the experimental findings on MBP denaturation. Panel b: Denaturation plot of MBP obtained by thermal equilibrium simulations at  $T = 0.75$ . The proper rescaling, (3.9), of RMSD makes simulation data collapse onto Ganesh *et al.* denaturation plot [118] (fraction of unfolded structures by circular dichroism spectroscopy).

such a minimal model context is the closest to the chemical one, as it reduces only the number of attractive non-bonded interactions responsible for the collapse onto compact native structures. That behavior mimics the chaotropic action of a solute (denaturant) which destabilizes the native states by competing with internal non-covalent protein interactions, such as hydrogen bonding and hydrophobic effect.

We thus expect that laboratory denaturation conditions can be effectively taken into account by a suitable choice of the cut-off radius. In fact panel *a* of Figure 3.5 depicts a contour plot of the Kabsch distance (mean root-square deviation of a structure with respect to the native state [106]) as a function of the simulation temperature and cut-off radius. The transition from the region where MBP structures are folded (black) to the denatured one (white) is sharp, thus pointing out the suitability of the selected approach to model chemical denaturation. Experimentally, Ganesh, Shah, Swaminathan, Surolia and Varadarajan have analyzed chemical MBP denaturation by guanidine hydrochloride (Gnd-HCl) at constant temperature,  $T = 28^\circ\text{C}$ . They have shown that the unfolded protein concentration  $\rho_u$ , at different GndHCl concentrations  $[D]$ , estimated via circular dichroism spectroscopy, could be fitted by means of a standard two-

state model (dashed line in Fig. 3.5 panel *b*). The latter Figure compares the denaturation of Gō-model MBP as a function of the cut-off radius  $R_c$ , with the denaturation curve in [118]. As a nativeness indicator for the numerical data we employ the already mentioned RMSD. Mapping between simulated and experimental denaturation results is established after a base-line subtraction of the RMSD values and a normalization to the  $[0, 1]$  interval. It reads

$$[D] = \frac{a}{R_c - R_c^0} - \frac{a}{R_c^* - R_c^0}, \quad (3.9)$$

where our reference interaction cut-off  $R_c^* = 7.5\text{\AA}$  corresponds to the fully native state  $[D] = 0$  in the experiment;  $R_c^0 = 4.0\text{\AA}$  is in turn the cut-off radius for complete denaturation and  $a$  is a tunable parameter adjusted to achieve data-collapse.

#### 4.2. Translocation dynamics

We characterize translocation success (top left panel of Fig. 3.6) in terms of translocation probability  $P_{\text{Tr}}$ . The fraction of translocated simulations is considered with respect to the total number of attempts, as a function of the importing force  $F$ . Figure 3.6 refers to  $T = 0.75$  and  $R_c = 6.8\text{\AA}$  that, as shown in Figure 3.5*b*, is the lowest  $R_c$ -value corresponding to native-like conformations, for both N-pulling (empty symbols) and C-pulling (filled symbols). In addition to  $P_{\text{Tr}}$ , the left side of Figure 3.6, *i.e.* panels *b* and *c*, provides also the fraction of *loss* events,  $P_L$ , and *unsuccessful* events  $P_U = 1 - P_L - P_{\text{Tr}}$ .  $P_L$  is the percentage of molecules that escape the capture by diffusion in the bulk.  $P_U$  is the fraction of proteins that are neither translocated nor *lost* within the time window  $T_w$ . The step-like shape of  $P_{\text{Tr}}$  vs.  $F$  allows a clear definition of the critical force  $F_c$  as the value for which  $P_{\text{Tr}}(F_c) = 1/2$ . If  $F \ll F_c$ , most of not-translocated runs are *lost*, since the importing force is so weak to be easily overcome by thermal motions. By increasing  $F$ ,  $P_L$  rapidly decays whereas the number of translocated and unsuccessful runs increases. For  $F \simeq F_c$ , almost all the untranslocated proteins correspond to unsuccessful runs and very few are *lost*. The overall scenario also indicates an asymmetric translocation process of folded MBP, which depends on the pulling terminus. The difference in  $F_c$  and in the number of proteins stuck in the pore between N- and C-pulling suggests that the transport pathways are different in the two cases.

To characterize the role of denaturation in the translocation mechanism, we estimate  $F_c$  at different cut-off radii. Panel *d* of Figure 3.6 shows that translocation at low denaturation (large  $R_c$ ) requires high forces. Decreasing  $R_c$ , the critical force is reduced and below  $R_c \simeq 6.8\text{\AA}$  it reaches a plateau. Actually, once denaturated ( $R_c < 6.8\text{\AA}$ , see Fig. 3.5), the dynamics of random coil MBP configurations becomes unaffected by a further reduction of  $R_c$ . In contrast, as for  $R_c > 6.8\text{\AA}$  the MBP structures become more and more compact and stable, the



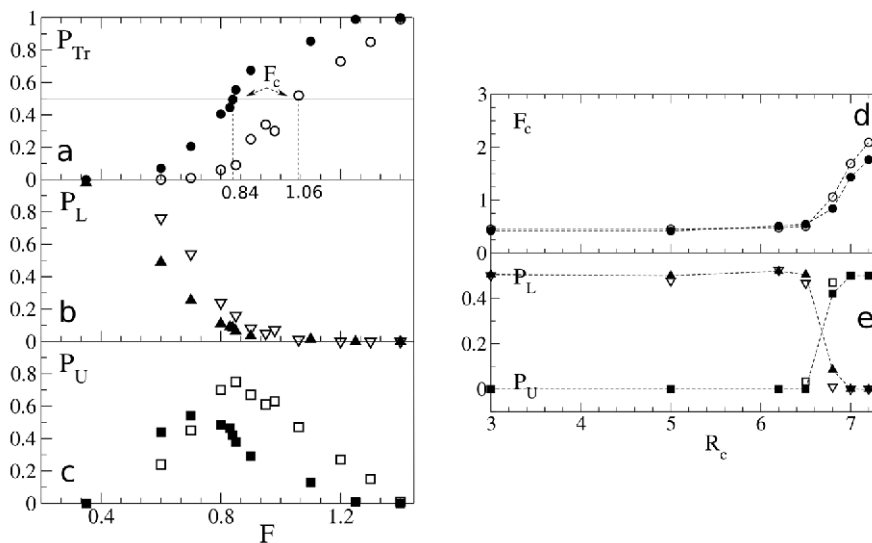


Figure 3.6: Here and in the following figures units are internal, *i.e.* related to the coarse-grained model we used. Distances, when present, are in Å. Empty symbols N-pulling, filled symbols C-pulling. Left hand side: Statistics of translocation as a function of the importing force  $F$  at  $R_c = 6.8\text{\AA}$  and  $T = 0.75$ : (a) translocation probability  $P_{Tr}$ , (b) loss event probability  $P_L$  and unsuccessful run probability  $P_U$  (c). Right hand side: Critical force  $F_c$  as a function of  $R_c$  for  $T = 0.75$ , panel d. Loss and *unsuccess* events probability,  $P_L$  and  $P_U$  respectively, at critical force, panel e.

increment of the critical  $F_c$  is naturally expected, due to a stronger resistance to unfolding. Therefore, the critical force is a quantity able to discriminate folded and unfolded MBP structures, in analogy to the applied voltage in experiments, (see [23, 27, 38]).

A further difference between denaturated and native-like MBP translocation is revealed by the amount of *loss* and *unsuccessful* events at critical force. Panel e of Figure 3.6 shows a clean transition at  $R_c = 6.8\text{\AA}$  in both  $P_L$  and  $P_U$  at critical force. At smaller  $R_c$  (denaturated state)  $P_L = 0.5$ , hence none of the unstructured chains gets stuck in the pore. That behavior is compatible with the idea of an importing force competing with thermal fluctuations to insert a residue in the pore:  $F_c \simeq F_{th} = k_b T / d_0 \simeq 0.2$ , with  $d_0 = 3.8\text{\AA}$  the average distance between consecutive residues. Once the first core of residues is imported, denaturated structures oppose a weak resistance and their translocation is easily accomplished. In contrast at larger  $R_c$ , the critical force is significantly greater than  $F_{th}$ . Hence, almost all MBP configurations start getting imported and those unable to finalize the transport within the allotted time end up stuck into the channel. At low denaturation, long pore blockades are entirely due to the structural resistance of folded MBPs to mechanical unfolding.

### 4.3. Residence time statistics

Figure 3.7 reports the average translocation (or blockade) time  $\tau_b$  as a function of the importing force for different values of the interaction cut-off  $R_c$ , for both the N-terminus and C-terminus pulling (panel *a* and *b* respectively). For a

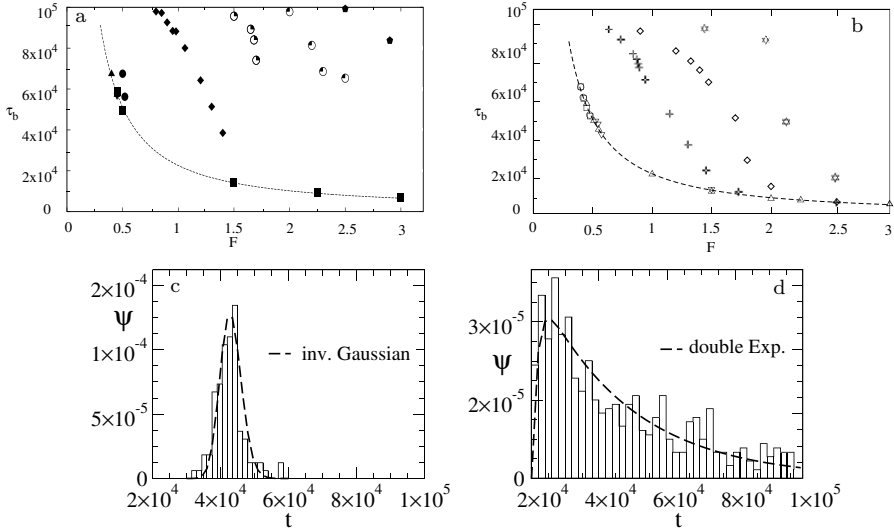


Figure 3.7: Panel a: Average translocation time on the sub-ensemble of successfully translocated runs  $\tau_b$  vs. importing force  $F$ . The different symbols concern different cut-off radii, namely  $R_c = 3.0, 5.0, 6.2, 6.5, 6.8, 7.0, 7, 2$  and  $7.5 \text{ \AA}$  are represented by upward triangles, downward triangles, squares, circles, diamonds, first-quadrant-filled circles, second-quadrant-filled circles and pentagons, respectively. The dashed line represents the fit (3.10). Data for N-pulling simulations. Panel b: Average translocation time on the sub-ensemble of fully translocated simulations  $\tau_b$  vs. importing force  $F$ . Different symbols represent different cut-off radii, namely  $R_c = \{3.0, 5.0, 6.2, 6.5, 6.8, 7.0\} \text{ \AA}$  and  $7.5 \text{ \AA}$ , squares, circles, downward triangles, upward triangles, cross, diamonds, stars respectively. The dashed line represents the fit, (3.10). Data for C-pulling simulations. Panels c and d: translocation time distribution for  $R_c = 6.5 \text{ \AA}$  and  $F = 0.575$  (denaturated) and  $R_c = 6.8 \text{ \AA}$  and  $F = 1.10$  (native) respectively. The dashed lines in panel c and d are an inverse Gaussian (3.12), and a double exponential fit (3.13).

given cut-off radius,  $\tau_b$  generally decreases with increasing  $F$ , as it is expected. However for  $R_c \leq 6.5 \text{ \AA}$ , the  $\tau_b$  vs.  $F$  data collapse onto a single curve (dashed), confirming that MBP structures are completely denaturated and react similarly to the importing force. The curve is well fitted by a three-parameter ( $\tau_0, F_0, \mu_0$ ) relation, namely

$$\tau_b(F) = \tau_0 e^{-F/F_0} + \frac{L}{\mu_0 F}, \quad (3.10)$$

which is a linear combination of an activation term  $\propto \exp(-F/F_0)$ , and a constant velocity drift  $\propto 1/F$ . The origin of the activation term can be explained as in what follows. An irreversible translocation can occur only after a stable core of residues is established inside the channel. This sort of capture process requires the overcoming of the initial entropic barrier due to the strong confinement of the chain induced by the channel, that even unstructured polypeptides experience [114]. For  $R_c \geq 6.8\text{\AA}$ , the average translocation time grows with the decrement of  $F$  (crosses, diamonds, stars in Fig. 3.7*a* and *b*), indicating that translocation is drastically slowed down. At large enough forces, overwhelming the stability of the native-like MBP,  $\tau_b$  tends to collapse on (3.10), regardless the value of the cut-off radius. The curve depicted by (3.10) thus constitutes a sort of base-line for all translocation times. In low-force regimes, the bending of data toward the horizontal line  $\tau_b = T_w$  is an effect due to the time window finiteness. In Figure 3.7*a* and *b*, that effect has been partially corrected by using the formula

$$\tau_b = \frac{1}{P_{\text{Tr}} + P_{\text{U}}} \left( P_{\text{Tr}} \tau_{\text{Tr}} + P_{\text{U}} T_w \right), \quad (3.11)$$

where  $\tau_{\text{Tr}}$  is the average time from the translocated runs and  $P_{\text{Tr}}$  is the corresponding probability (see Fig. 3.6). The use of (3.11) can be justified. Let us consider a set of  $M$  runs, where  $M_{\text{T}}$  of them translocate within  $T_w$ , while  $M_{\text{U}}$  do not accomplish the passage but are expected to translocate in a time greater than  $T_w$  and  $M_{\text{L}}$  are lost (will practically never translocate), obviously  $M = M_{\text{T}} + M_{\text{L}} + M_{\text{U}}$ . By definition, after discarding lost events, the average translocation time is

$$\tau_b = \frac{1}{M_{\text{T}} + M_{\text{U}}} \sum_{r=1}^{M_{\text{T}}+M_{\text{U}}} t_r,$$

where  $t_r$  is the translocation time of the  $r$ -th run (thought for  $T_w \rightarrow \infty$ ). We could rewrite  $\tau_b$  by separating the runs for which  $t_r < T_w$  from the runs for which  $t_r \geq T_w$  and, consequently, split the summation into

$$\tau_b = \frac{M_{\text{T}}}{M_{\text{T}} + M_{\text{U}}} \frac{1}{M_{\text{T}}} \sum_{r=1}^{M_{\text{T}}} t'_r + \frac{M_{\text{U}}}{M_{\text{T}} + M_{\text{U}}} \frac{1}{M_{\text{U}}} \sum_{r=1}^{M_{\text{U}}} t''_r.$$

Here, the first average involves the terms  $t'_r < T_w$ , while the second one the terms  $t''_r \geq T_w$ . By considering that  $P_{\text{U}} = M_{\text{U}}/M$  and  $P_{\text{Tr}} = M_{\text{T}}/M$ , we get the expression

$$\tau_b = \frac{1}{P_{\text{Tr}} + P_{\text{U}}} \left( P_{\text{Tr}} \tau_{\text{Tr}} + P_{\text{U}} T_w \right),$$

where we have used the lower bound for  $t''_r \geq T_w$  and the formula  $\tau_{\text{Tr}} = \frac{1}{M_{\text{T}}} \sum_{r=1}^{M_{\text{T}}} t'_r$ .

Panel *c* of Figure 3.7 shows a typical translocation time distribution  $\psi(t)$  of a denaturated MBP. It looks well localized around its average  $\tau_b$  and it is properly fitted by an inverse Gaussian, with parameters  $D_0, \mu_0$ :

$$\psi(t) = \frac{L}{\sqrt{4\pi D_0 t^3}} \exp\left\{-\frac{(L - \mu_0 F t)^2}{4D_0 t}\right\}, \quad (3.12)$$

that is the first-passage time distribution of biased random walkers on the interval  $[-\infty, L]$  emitted in 0 and absorbed at  $-\infty$  and  $L$  [57]. An instance of the time distribution  $\psi(t)$  of native-like MBP translocation is plotted in panel *d*. In this case,  $\psi(t)$  has a fatter large-time tail and it is well reproduced by a double exponential:

$$\psi(t) = \frac{k_1 k_2}{k_2 - k_1} \left[ e^{-k_1(t-t_0)} - e^{-k_2(t-t_0)} \right], \quad (3.13)$$

with rates  $k_1, k_2$  and  $t_0$  being an offset time interpreted as the time taken by a denaturated MBP to cross the pore. As we shall see below, the double exponential is justified by the presence of two successive stall points, corresponding to two energy barriers, the overcoming of which can be seen as an activated processes. Translocation phenomenology is better characterized by addressing the

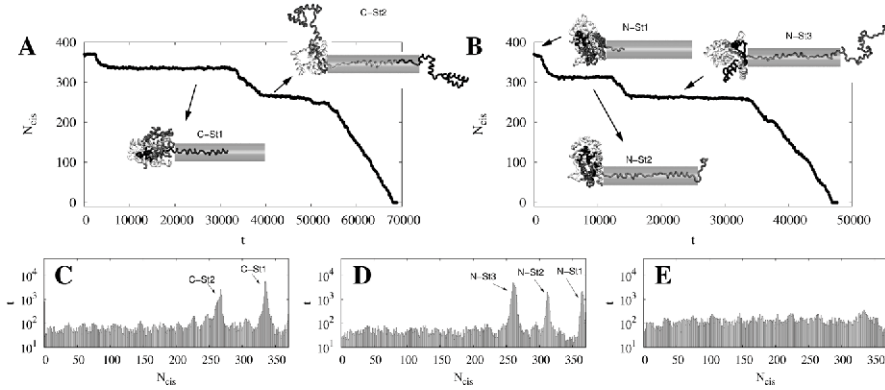


Figure 3.8: Time evolution of the number of residues on the cis-side of the pore  $N_{cis}$  for a pulling run at  $R_c = 6.8 \text{ \AA}$  and critical force (panel A: C-terminus pulling, panel B: N-terminus pulling). Snapshots of representative configurations are reported for the plateaus corresponding to the protein blockades. The average time spent by the protein in different  $N_{cis}$ , at critical force, is reported in panels C, D and E for  $R_c = 6.8 \text{ \AA}$  C-pulling,  $R_c = 6.8 \text{ \AA}$  N-pulling,  $R_c = 6.5 \text{ \AA}$  C-pulling, respectively.

time evolution of the number of residues  $N_{cis}$  on the cis-side of the pore. Panel A of Figure 3.8 shows  $N_{cis}(t)$  for one successful run at critical force for  $R_c = 6.8 \text{ \AA}$ , pulled from the C-terminus. Clearly, most part of the time is spent by the MBP in two particular stalling states,  $N_{cis} \approx 335$  (C-St1) and  $N_{cis} \approx 267$  (C-St2), corre-

sponding to configurations where either residue 335 or residue 267 are respectively located at the pore entrance. Representative snapshots of the two states are included in the figure for illustration. To gain more quantitative information we compute the time  $t_r(N_{cis})$  spent by the  $r$ -th run in a given  $N_{cis}$  state during translocation. Time  $\bar{t}(N_{cis})$ , defined as the average of  $t_r(N_{cis})$  over the ensemble of translocated runs, is plotted in panel C of Figure 3.8, for C-pulling translocations at critical force and  $R_c = 6.8\text{\AA}$ . The two peaks in the histogram correspond to the two blockade events shown in panel A. Clearly, translocation is far from being uniform and looks more as a step-like (or *stick and slip*) process, where the protein has to overcome successive free-energy barriers associated with specific structural rearrangements. The analysis of N-pulling (same cut-off radius and corresponding critical force), panels B and D in Figure 3.8, confirms the overall picture: also in this case the protein spends most part of the time in specific configurations. The stalling events are three and take place at different positions, namely  $N_{cis} = 363$  (N-St1), corresponding to residue 7 at the pore mouth,  $N_{cis} = 307$  (N-St2, residue 63 at the mouth) and  $N_{cis} = 256$  (residue 114, N-St3). To complete the phenomenology, we note that the transport of denaturated proteins ( $R_c < 6.8\text{\AA}$ ) appears much more uniform (peak-less), panel E of Figure 3.8.

The scenario for both N and C-pulling is robust under changes of both  $R_c$  (provided that  $R_c \geq 6.8\text{\AA}$ , *i.e.* native-like MBP) and pulling force range  $F$ , with only slight differences in the peak intensity (Fig. 3.9). Concerning the *unsuccessful* translocations, we find that the already mentioned stall points still occur (Fig. 3.9 and Fig. 3.10), just to confirm their leading role in transport bottlenecks. In summary, panel B of Figure 3.1 sketches in a 2D-topological view the positions of the five stall points detected for structured protein translocation runs (residues 7, 63, 114, 267, 335).

The loss events can provide insights into the translocation process and might be related to experimental observations. A typical event observed in loss simulations is that the protein spends some time at the pore entrance, attempting to insert the terminus, before definitely escaping the capture by diffusion. The duration of the events in which the number of residues inside the pore is greater than 0 is well determined in lost simulations and defines a sort of protein-pore interaction time  $t_I$ . Moreover, reasonably assuming that in between two successive events the protein remains close enough to the pore to generate some kind of blockage, it is possible to attempt a closer connection between simulation results and experimental findings. Panel A of Figure 3.11 depicts the interaction time distribution obtained for a denaturated protein while panel B shows the trend of the average interaction time  $\tau_I$  as a function of the denaturation degree (as usual characterized by the Gō-model interaction cut-off  $R_c$ ) for different importing force intensities. A common behavior of unfolded structures is evident, while the average interaction time for compact molecules results greater and dependent on the cut-off radius. Although a quantitative comparison might

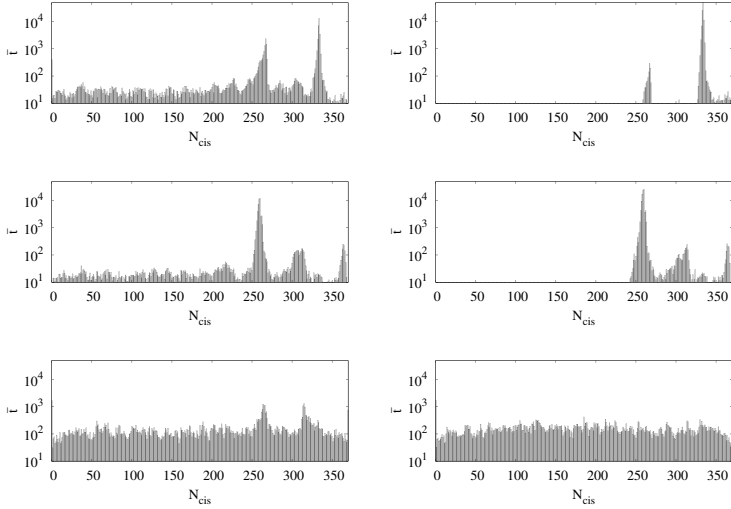


Figure 3.9: Average time spent by the protein as a function of  $N_{cis}$ . Panels of the first row refer to translocated simulations (left), and unsuccessful runs (right), both for C-pulling at critical force,  $R_c = 7.2\text{\AA}$ . The latter panel clearly shows the presence of the two main blockades (C-St1 and C-St2), which can be detected also for successful simulations, thus confirming the robustness of the data shown in panel C of Figure 3.8 with respect to changes in  $R_c$ . In the second row, analogous results are reported for N-pulling translocations: three main peaks are evident (N-St1, N-St2 and N-St3). Finally, in the last row the same data for successful simulations for  $R_c = 6.5\text{\AA}$  (left) and  $R_c = 6.2\text{\AA}$  (right), both for N-pulling at critical force are depicted. For  $R_c = 6.5\text{\AA}$  the peak structure still partially holds (two on three peaks are still detectable) and it is necessary to decrease the value of the cut-off radius to  $6.2\text{\AA}$  to recover a homogeneous behavior equivalent to the one shown in panel E of Fig 3.8, where the pulling is from the C-terminus.

be too ambitious, the scenario we observe in the simulations closely resembles the presence of the ‘bumping events’ reported in the experiments by Mestorf *et al.* [28] and classified as short-time low-intensity current blockades. In practice, the protein moves close to the pore partly blocking the current but does not start a translocation. The bumping time distribution results to be well fitted by an exponential (Fig. 3.11A) and its average is about one order of magnitude lower than the corresponding average translocation time, a fact qualitatively in agreement with the experimental data (for the numerical specific case here considered  $\tau_b/\tau_I \simeq 50$ ).

#### 4.4. Stretching vs. translocation

The previous picture of MBP translocation rises the natural question as to why the transport of initially compact structures becomes temporarily stalled at well defined stages. We relate these ‘rate limiting steps’ of translocation to

## Coarse-grained molecular dynamics and continuum models

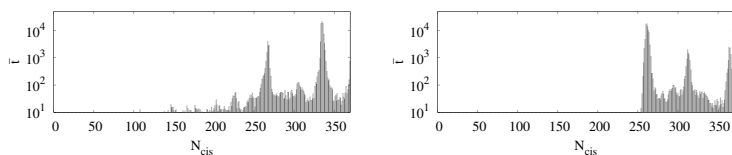


Figure 3.10: Average time spent by the protein as a function of  $N_{cis}$  for unsuccessful runs for  $R_c = 6.8 \text{ \AA}$ . Left panel refers to C-pulling, right panel to N-pulling. The two peaks in the left panel and the three in the right one correspond to the peaks in panels C and D of Fig 3.8, respectively.

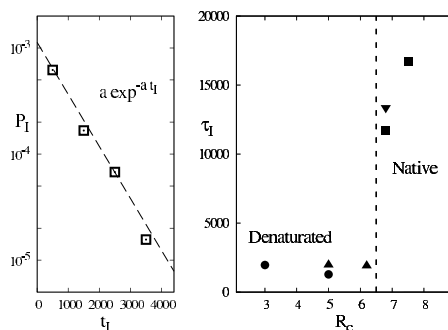


Figure 3.11: Panel A: Protein-pore interaction time ( $t_I$ ) distribution and associated exponential fit for unfolded protein and small pulling force. Panel B: Trend of average interference time as a function of the pulling force and cut-off radius. Unfolded structures show a common behavior at short  $\tau_I$  while a greater average interaction time is associated to compact structures. The vertical dashed line marks the  $R_c$ -threshold separating in the Gō-model approach native-like from fully denaturated conformations. Different symbols correspond to four increasing force intensities (in the order: squares, circles, up and down triangles).

the MBP structural properties that govern the free-space stretching. As mentioned in the preamble of this chapter, targeted AFM experiments [98] have shown that MBP mechanical stretching occurs via a sequence of events corresponding to the successive breakdown of specific domains, termed unfoldons. The analogy of pulled translocation with AFM mechanical stretching suggests that unfoldons might be involved also in MBP translocation bottlenecks. Preliminarily, we check, via standard Steered Molecular Dynamics stretching protocol [117] that the Gō-model of the MBP is able to reproduce the ‘unfoldon picture’ observed in the experiments [98]. In Figure 3.12A the average over 50 runs of the force-extension (end-to-end distance  $R_{ee}$ ) curve of the MBP in the numerical stretching experiment is depicted. The three branches, denoted by M1, M2+M3 and M4 according to the nomenclature used in [98], separated by the three worm-like-chain curves [119], identify the same unfoldon opening sequence described in the reference experimental work, including the coupled un-

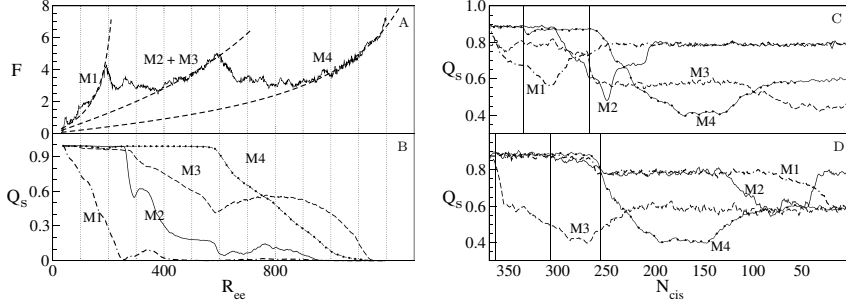


Figure 3.12: Right: force (A) and average value of fraction of native contacts  $Q_s$  (Mk)  $k = 1, 4$  (B) vs. the end to end distance  $R_{ee}$  in stretching simulation at pulling velocity  $v = 0.05$ , temperature  $T = 0.5$  and cut-off radius  $R_c = 7.5\text{\AA}$ . Dotted lines represent worm-like chain fits [119]. Left: average value of  $Q_s$  (Mk) as a function of  $N_{cis}$  for  $R_c = 6.8$  and  $T = 0.75$  for N-pulling (C) and C-pulling (D) translocated runs at critical force. Vertical black lines mark the residues mainly involved in the translocation bottlenecks.

folding of M2 and M3. Also, the stretching force required to perform constant velocity detachment follows a peculiar trend qualitatively in compliance with the experimental evidence (see Fig. 3.13 for a quick summary of the latter). The picture becomes clear when plotting the average fraction of the active native contacts  $Q_s(Mk)$ ,  $k = 1, \dots, 4$  in each unfoldon as a function of the end-to-end distance  $R_{ee}$  of the MBP. For instance, in the initial stage of the elongation, corresponding to the curve branch labeled by M1, the structural parameter  $Q_s(M1)$ , dash-dotted line, decreases indicating the rôle of M1 detachment in this stage. The simultaneous opening of M2 and M3, solid and dashed line respectively, is also evident in the second branch. Finally M4 opens. Snapshots of the detachment sequence are shown in Figure 3.14.

To understand the role of the unfoldons in the MBP translocation process, first we locate the five blockage-points with respect to the unfoldons in the protein structure (Fig. 3.1). Then, again, we plot the unfoldon structural parameters  $Q_s(Mk)$  as a function of  $N_{cis}$  for C-pulled and N-pulled translocations, respectively (Fig. 3.12C and D). In the C-terminus pulling case (Fig. 3.12C), the sequence of unfoldon opening is essentially the same as in the mechanical stretching. First, M1 breaks down, followed by M2 and M3, which dynamics is again correlated, and finally the M4-opening concludes the process. The vertical dotted lines in Figure 3.12C and D highlight the two stall points C-St1, C-St2 in the C-pulling and the three other ones, namely N-St1, N-St2, N-St3, in the N-pulling (see discussion of Fig. 3.8). A very weak correlation emerges between unfoldon dynamics and blockades. Specifically, the two configurations reported in panel A of Figure 3.8 (see also Fig. 3.1B and the vertical lines in Fig. 3.12C and D) show that the first blockage-point for C-pulling (residue 267) takes places in unfoldon M1 (darkest gray in the figure), while the second one (residue 335) lies in the initial part of M2 (light gray). In the pulling from the N-terminus, we observe



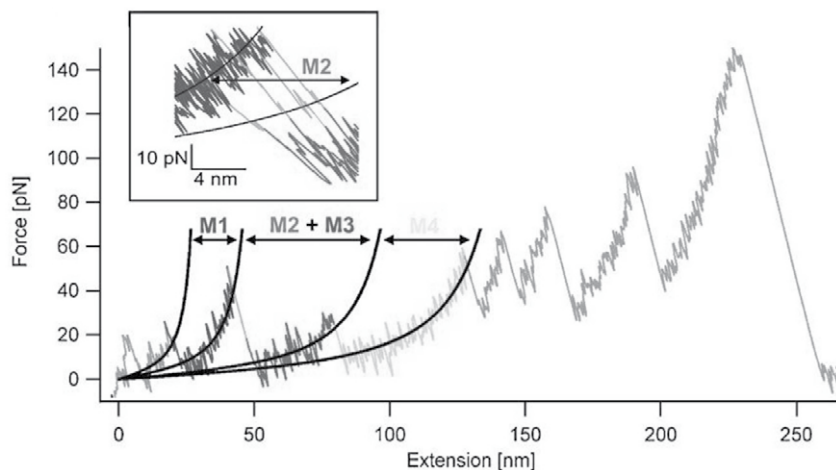


Figure 3.13: Adapted from [98]. Characteristic force-extension trace of a ddFLN1-5-MBP chimera [98]. Unfolding of the native protein and subsequent elongation of the polymer is shown in the darkest gray (unfolding M1), the breakdown of M2 is shown in light gray, the detachment of M3 is shown in dark gray and the unfolding of M4 is shown in the lightest gray. Contour length increases are marked on worm-like-chain (WLC) curves fit to the data. At extensions  $> 125\text{nm}$  the characteristic unfolding pattern of three ddFLN scaffold domains is visible (gray, leftmost and rightmost ends), followed by detachment of the protein. Inset: Superposition of several MBP unfolding traces zoomed into the region of the short-lived intermediate state M34 (dark gray) resulting from the detachment of M2 (light gray). A best WLC fit to the curves is shown in black. All light gray levels (lifetime of the intermediate state M34) fall on the same WLC curve, indicating the breakdown of a well-defined state.

that M3 is the first to be broken followed by M4, then M2 and finally M1. Again the blockade-points do not appear directly related to the unfolding boundaries, panel B of Figure 3.8, with the sole exception of the stall N-St3, located at the boundary between M3 and M4.

## 5. Discussion

An essential result of our analysis is a sharp change in the translocation dynamics upon varying the cut-off radius from  $R_c = 6.5\text{\AA}$  to  $R_c = 6.8\text{\AA}$ . We recall that, according to (3.9), a mapping between  $R_c$  and the MBP denaturation degree can be established (Fig. 3.5).

**Denaturated-MBP translocation** ( $R_c \leq 6.5\text{\AA}$ ). Below  $R_c = 6.5\text{\AA}$ , the MBP behaves as a random coil: its translocation dynamics is basically independent of  $R_c$  as revealed by the critical force behavior (Fig. 3.6d) that remains almost constant for  $R_c \leq 6.5\text{\AA}$  and by the blockade time data (Fig. 3.7a and b) that

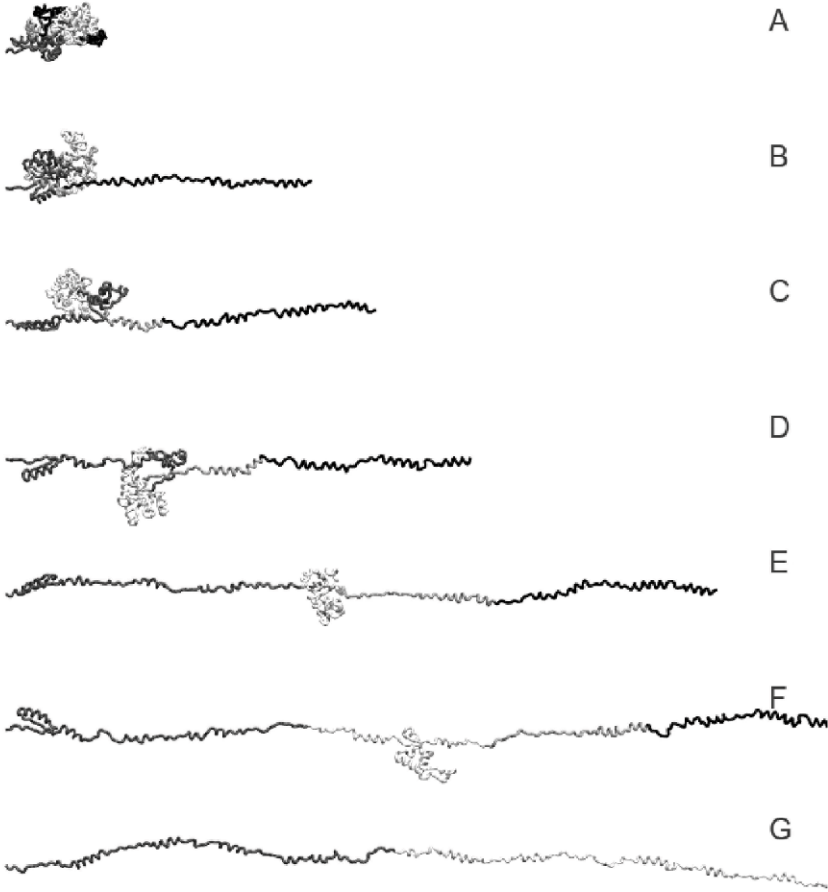


Figure 3.14: Snapshots of MBP unfolding at  $T = 0.5$ ,  $R_c = 7.5 \text{ \AA}$  and  $v = 0.001$ . The unfolding sequence is the one described by Bertz and Rief, [98]. The first part to open is the unfoldon M1 (darkest gray) as shown in panel B, then M2 and M3 unfold (light and dark gray), panels C, D and E. With the exception of the initial portion of M3, the last structure to detach is M4 (lightest gray), panels F and G. In panel, A the initial folded MBP conformation is depicted.

collapse onto the same curve for every  $R_c$ , thus exhibiting a common behavior, (3.10). In this  $R_c$ -range, not only the critical force  $F_c$  but also the probability of lost events at  $F_c$  shows a plateau (Fig. 3.6e). Moreover, the pulling terminus (C or N) does not influence significantly any of these quantities. The sole barrier in the translocation dynamics is due to the pore entrance; once the protein is captured, the average time spent at different stages of the process is similar (panel E of Fig. 3.8). Thus, random-coil MBP translocation can be interpreted as a capture (activated) stage followed by an elementary first passage process, where random

walkers under a constant bias are injected from the cis-side of the channel and absorbed to the trans-side.

**Native-like MBP translocation** ( $R_c \geq 6.8\text{\AA}$ ). Folded structures exhibit a much more complex dynamics, characterized by a transport getting stalled in the pore due to specific MBP configurations. Translocation times are greater than the unfolded case. Their average increases with  $R_c$  (Fig. 3.7a and b) and their distributions show fat tails (Fig. 3.7d). These results agree with the experiments described in [23], where, in the explored forcing regime, only denatured proteins were able to rapidly translocate, whereas very long blockades pertained to partially folded ones. The capability of numerical techniques to explore the dynamical details evidences the presence of stable and reproducible bottlenecks of the transport, namely C-St1, C-St2, N-St1, N-St2, N-St3 in native-like MBP transport (Figs. 3.8, 3.9 and 3.10). Natural candidates for such stall points are the boundaries between MBP unfoldons. However, our results clearly indicates that the AFM-unfolding dynamics is not correlated to the stalls.

Rather, the analysis of MBP native-contact maps allows us to shed light on the actual mechanism underlying the stalls. Indeed, the blocking C-St1 and C-St2, but also their counterparts for the N-pull case, can be interpreted in terms of a specific sequence of native-contact breaking an MBP sub-domain undertakes in order to engage the pore in an almost linear conformation. In fact, the *stalling domains* of the protein are those with the greater amount of *external native contacts*, where external means ‘excluding all the inner contacts of the domain’. In formulas, if  $K$  is the domain, then its contact density reads

$$\rho_c(K) = \frac{1}{m_K} \sum_{i \in K} \sum_{j \notin K} M_{ij}, \quad (3.14)$$

where  $m_K$  denotes the number of residues in  $K$  and  $M_{ij} = 1$ , if  $i$  and  $j$  form a native contact, and  $M_{ij} = 0$  otherwise. In Figure 3.15 the contact map of native MBP, obtained for  $R_c = 6.8\text{\AA}$  is considered. The lower side panel shows the average of external contacts composed by the ten-residue-long consecutive regions of the MBP. Peaks identify the protein regions with the greater number of contacts that are thus the most probable candidates to cause the stall points once engaging the pore entrance. However, by considering the dynamics data (like the configurations associated with the stalls, Fig. 3.8C for example), we can be more precise by selecting the two critical regions around the peaks C-St1 and C-St2: C-K1 = [328, 338] and C-K2 = [260, 270], respectively, and the complementary regions X1 = [339, 370], X2 = [271, 327] and X3 = [1, 259], so to patch the whole molecular structure. Table 3.1 summarizes the density of external contacts (see (3.14)) that these regions form with the part of the molecule running from the proper boundary of the issued region to the free terminus (a generalization of this approach concerning only a static analysis of contact maps - without resorting to dynamical data - is provided in the next chapter). For instance, ex-

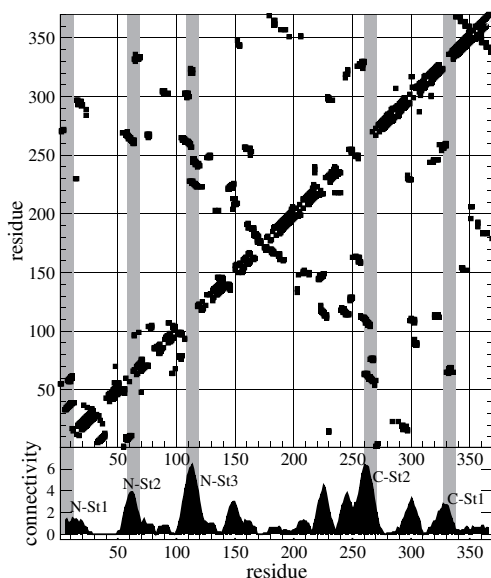


Figure 3.15: Contact map of native MBP obtained for  $R_c = 6.8\text{\AA}$ . Bottom panel depicts the average number of long range contacts each residue establishes. The peaks identify the regions of the chains with the larger density of external contacts, which are responsible for putative stall points. Vertical gray bands highlight the stalling regions as obtained by the analysis relevant to Figure 3.8. These regions form a significant number of contacts with distal parts of the molecule and only when the contact cluster at the pore cis-side mouth is broken the translocation can proceed further on, up to the next bottleneck (next cluster).

ternal contacts pertaining to C-K1 bond this domain with residues from 327 to 1 (N-Terminus), as here C-terminus pulling blockades are exemplified. This approach allows tracking only the effective external contacts that come into play when a region is facing the pore, namely the contacts formed with just that part of the molecule which has not yet engaged the pore (*backward contacts*). The higher values pertain to C-K1 and C-K2 sub-domains. They confirm the picture emerging from Figure 3.15. These areas, being with high contact density, oppose the maximal resistance to unfolding once in front of the pore. In other words, translocation bottlenecks are determined by those sub-domains that, still partially folded when approaching the pore cis-side, carry with themselves other distal regions of the molecule tightly bonded to by native interactions. The contact analysis leads to the same conclusion also for the N-terminus pulling, where the critical regions are  $N\text{-K1} = [1, 11]$ ,  $N\text{-K2} = [55, 63]$ ,  $N\text{-K3} = [105, 115]$ .

C-pulling			N-pulling		
Name	Region	$\rho$	Name	Region	$\rho$
X1	[339, 370]	0.75	N-K1	[1, 11]	3.64
C-K1	[328, 338]	2.00	Y1	[12, 54]	0.47
X2	[271, 327]	1.07	N-K2	[55, 63]	2.33
C-K2	[260, 270]	5.09	Y2	[64, 104]	0.93
X3	[1, 259]	0.00	N-K3	[105, 115]	5.45
---	-----	---	Y3	[116, 370]	0.00

Table 3.1: Contact density of MBP subdomains described in the text. Col.1: domain nomenclature, C-K1, C-K2 and N-K1, N-K2, N-K3 stand for critical blocks in C- and N-pulling respectively, X1, X2, X3 and Y1, Y2 and Y3 are the complementary regions. Col.2: residues involved in each sub-domain. Col.3: density of external native contacts as defined in (3.14). Critical blocks have a contact density larger than complementary regions.

## 6. Remarks

We have simulated Maltose Binding protein translocation across  $\alpha$ HL nanopore via a coarse-grained computational model for both the MBP and the pore. As the channel is narrow, translocation properties strongly depend on the denaturation state of the MBP. In the issued Gō-like model, molecule denaturation is controlled by the parameter  $R_c$  determining the number of native attractive interactions. In the region ( $6.5 < R_c < 6.8$ )Å, a transition is observed from random-coil MBP (denaturated) to native-like structures. The transition emerges from both equilibrium (Fig. 3.5) and transport simulations (Fig. 3.6). In particular, translocation of denaturated MBP is almost uniform and consists of a capture stage followed by a simple driven diffusion process. The passage in the channel of folded MBP is more critical and interesting, it looks like a *stick-slip* dynamics characterized by constant-velocity transport broken by stalling events in the channel (Fig. 3.8). For instance, the C-terminus translocation occurs via two long stall events resulting in a double-exponential tail-behavior of the translocation time distribution (Fig. 3.7d). This behavior is presumably associated with the presence of two subsequent free-energy barriers that MBP has to overcome in order to complete the passage. Our analysis also shows that stall events are related to the MBP regions with high density of native external contacts. Thus, in principle, long blockade events and stall points can be predicted by looking directly at the MBP PDB-structure. In contrast, a weak correlation is found between stall points and unfoldons, the structural blocks through which the MBP reacts to mechanical stretching in free space. This result is a strong indication that, despite the analogy between pulled translocation and mechanical unfolding, the pathways gathered from mechanical pulling are not sufficient to make inferences on translocation

mechanisms. The action of the pore, indeed, drastically modifies the unfolding pathway during the transport with respect to a free pulling process.



# Chapter 4

## 1-Dimensional Continuum Models

### 1. Preamble

In order to interpret numerical results it is appropriate to build suited analytical models of the physical and chemical processes that occur in real translocations and to define proper statistical observables to grasp and summarize the essence of the phenomenon.

The first section of the present chapter is dedicated to an overview of the essential ideas behind low-dimensional continuum models used in the literature to summarize both experimental and three-dimensional molecular dynamics results.

Section 3 highlights the practical implementation of the specific approach performed to obtain the free energy profile for the translocation simulation protocol described in chapter 3. Therefore, umbrella sampling simulations and multiple weighted histogram analysis method on coarse-grained molecular dynamics (MD) simulations are outlined [96]. The proteins under scrutiny are again described by the Gō-like model already introduced, which accounts accurately for the influence of the native structures on both folding and translocation pathways [56, 114, 115, 120]. Finally, section 4 expounds the achieved results and presents a comparison with the 3D numerical investigations. Also, since a mutant version of the MBP (termed MalE262) and an additional protein, the Endoglucanase Cel5A (PDB-ID:1a3h), have been considered in this perspective, relevant outcomes are provided as well.

### 2. Overview

#### 2.1. The general framework

Experimental results highlight the presence of energy barriers that the molecules have to overcome to accomplish translocation [121]. Also, empirical distributions of blockade times show tail asymmetry typical of phenomena where the motion is due to diffusion and drift effects [16, 21]. However, even for a coarse-grained description of the protein dynamics, the energy landscape is still too high-dimensional to allow a concise representation of the process.



It is hence convenient to project the system evolution over relatively few effective coordinates,  $Q_1, \dots, Q_k$  (termed *reaction coordinates* or *collective variables*), functions of the positions of the amino acids constituting the molecule [122]. A translocation coordinate should enable us to parametrize the walk of the protein along the pore. Since the evolution of each  $Q_j$  is expected to undergo random fluctuations, we are interested in the probability,  $P(Q_1, \dots, Q_k, t)$ , for the system to visit a state  $Q_1, \dots, Q_k$  at time  $t$ , given an initial condition  $P(Q_1, \dots, Q_k, 0)$ . Two main questions arise: i) What are the most representative reaction coordinates  $Q_j$  in order to properly describe a translocation process? ii) What is the equation governing the evolution of  $P$ ?

In biopolymer translocation, a set of reaction coordinates that we can consider appropriate should smoothly trace without ambiguity the pathways between the two states (*cis* and *trans*) with respect to the pore ends. The common cylindrical symmetry of the narrow pores suggests to project the chain dynamics along the channel axis ( $x$ -axis without loss of generality) and consider nugatory the transversal motion ( $y, z$ ). Hence, the kinetics can be reasonably parametrized by a single reaction coordinate  $Q$ , expected to be a function of the  $x$ -coordinates of the residues only. The choice of  $Q$  is not a trivial task, there is neither a unique option nor a general rule, and usually an appropriate choice depends on the specific problem. For instance, in the case of a small protein pulled through a semi-infinite [54] or long enough [55] pore, the natural reaction coordinate can be either the  $x$ -position of the pulled residue or the  $x$ -component of the center of mass. Other authors found it reasonable to introduce, as reaction coordinate, the number of monomers in a given region of the space [74, 79, 123].

The derivation of an evolution equation for  $P$ , starting from the equations of motion of the full system, is generally problematic. Moreover, in general, the dynamics of  $Q$  in time is a stochastic process. A simplified approach is based on the assumption that the evolution of the collective variable is the slowest in the system, with a strong separation from other time scales, supposed irrelevant. Translocation is then considered completed when the stochastic process  $Q(t)$  crosses for the first time a threshold value  $Q_{\text{th}}$ , corresponding to the protein exiting the pore from the trans-side. This point of view clearly defines a first-passage problem in a time window  $T_w$ , with exit time  $t_{\text{out}}$  defined by

$$t_{\text{out}} = \min_t \{0 \leq t \leq T_w \mid Q(t) = Q_{\text{th}}\}$$

In this framework, the translocation phenomenology can be conveniently recast in the motion of an effective particle at place  $Q$ , undergoing a driven diffusion in a potential of mean force  $V(Q)$  [57]. Here,  $V(Q, t) = G(Q) + W(Q)$  and  $G(Q)$  is the free-energy profile in the variable  $Q$  of the unperturbed system, while  $W(Q)$  is the work done by the importing force. In a first approximation,

we can presume that the probability  $P(Q, t)$  obeys the Smoluchowski equation

$$\frac{\partial P}{\partial t} + \frac{\partial J}{\partial Q} = 0, \quad (4.1)$$

$$J(Q, t) = -\mu_0 P \frac{dV(Q)}{dQ} - D_0 \frac{\partial P}{\partial Q}, \quad (4.2)$$

where  $\mu_0$  is the effective mobility coefficient and  $D_0$  the effective diffusion one. At equilibrium, they satisfy the fluctuation-dissipation relation:  $\mu_0 = D_0/(k_b T)$ . In (4.1), the contributions of the fast degrees of freedom are collected in a single diffusion term  $\partial(D_0 \partial P / \partial Q) / \partial Q$ . The Smoluchowski equation is satisfied by the probability to find the reaction coordinate at place  $Q$ , at time  $t$ . Such an equation pertains, thus, to a population of random walkers undergoing an overdamped Langevin driven diffusion in the potential of the mean force  $V(Q)$ . As far as the results stemming from this approach are concerned (sec. 4), (4.1) has not been solved directly, rather we have found it more convenient to solve several 1D Langevin runs from (4.3), an approach that results equivalent to solving the Smoluchowski equation.

$$\dot{Q}(x) = \frac{-\nabla_x V(Q(x)) + \Gamma(t)}{\gamma} \quad (4.3)$$

$$V(Q(x)) = G(Q(x)) + W(Q(x))$$

In stochastic-process first-passage theory (see [57]), the basic quantity is the survival probability  $S(t)$ , given by

$$S(t) = \int_{Q_0}^{Q_{th}} dQ P(Q, t), \quad (4.4)$$

that estimates the number of systems for which  $Q$  is still in the interval  $[Q_0, Q_{th}]$  as a function of time, corresponding to the protein still in the channel. Here  $Q_0$  is the value of  $Q$  at the pore entrance. Hence,  $1 - S(t)$  amounts to the probability to be outside the pore at time  $t$ . Therefore its derivative

$$\psi(t) := -\frac{dS(t)}{dt} = \int_{Q_0}^{Q_{th}} dQ \frac{\partial P(Q, t)}{\partial t}$$

determines the distribution  $\psi(t)$  of the exit times. Thus, by using the Smoluchowski (4.1), we get:

$$\psi(t) = J(Q_{th}, t) - J(Q_0, t), \quad (4.5)$$

namely the exit-time distribution is fully determined by the values of the probability current  $J(Q, t)$  at the boundaries (pore ends). Common choices

for a generic boundary condition  $Q_B$  can be the following ones: absorbing  $P(Q_B, t) = 0$ , reflecting  $J(Q_B, t) = 0$  or mixed  $J(Q_B, t) = K_B P(Q_B, t)$ . The last one seems to be more realistic for describing translocation processes, as real pore ends cannot be considered neither perfectly reflecting nor perfectly absorbing. The solution  $P(Q, t)$  to (4.1) cannot be found in general for an arbitrary potential. However, even without solving the Smoluchowski equation, the exact analytic expression can be found for the average translocation time  $\tau_b(F, L)$  and probability  $P_{\text{Tr}}(F, L)$  as a function of the importing force  $F$  and the pore length  $L$  [55, 76, 77]. The theoretical prediction of  $\psi(t)$  requires instead a complete closed solution. Therefore, only simple cases can be worked out or approximated expressions can be obtained [55, 75, 77, 114] and, in specific conditions, we can reproduce both simulation and experimental data. For instance, the inverse Gaussian function (3.12) is the first-passage time distribution of a diffusion process in the interval  $[-\infty, L]$ , starting from  $x = 0$ , with absorbing boundaries in  $-\infty$  and  $L$ , and driven by a constant drift  $\mu_0 F$  [57]. Despite its simple origin, the approach is able to match observed translocation time distributions under specific conditions [38].

In [55] the phenomenology of the transport appears more complicated than a simple driven diffusion, but the mild shape of the free-energy (Fig. 2.6) permits a simple interpretation of the translocation kinetics in terms of a single jump over an extended barrier followed by a diffusion driven by a constant force. The theoretical predictions in [55] were corroborated by translocation time histograms (Fig. 4.1) which are well fitted by first-passage-time distributions  $\psi(t)$ , obtained by numerically solving the Smoluchowski equation (4.1) in the spatial coordinate  $x$  in  $[0, L]$ . Only a constant-force drift has been used in [55] under mixed boundary conditions, namely  $J(0, t) = -K_0 P(0, t)$ ,  $J(L, t) = K_L P(L, t)$ , where  $K_0$  and  $K_L$  have been considered adjustable parameters.

It appears then necessary to calculate the free energy profile of the whole process to take into account the barriers that hamper the crossing. A quite common approach for similar issues refers to the umbrella sampling method [95], a computational technique used to improve sampling of systems which phase space is divided into subdomains separated by high energy jumps. The umbrella sampling technique is based on the choice of a fictitious potential to overcome the energy constraints so to explore even the most unlikely areas of the phase space. The thermodynamics properties calculated in this way need to be deweighted from the influence of the artificial potential and recombined in order to provide the unbiased probability (free-energy). This purpose can be achieved, for example, by employing the weighted histogram analysis method (WHAM) [96, 124].

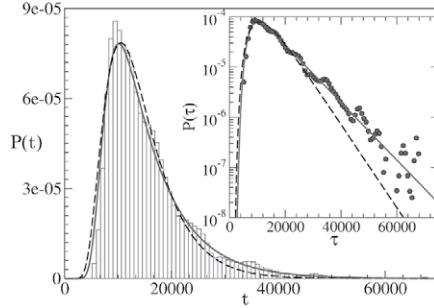


Figure 4.1: Adapted from [55]. Probability distribution of translocation times for Ubiquitin translocation in a  $L = 300\text{\AA}$  pore [55]. The dashed line is an inverse Gaussian fit (3.12) of the histogram while the continuous line is the results obtained solving (4.1) in the domain  $[0, L]$  with  $V(x)$  from Fig. 2.6, constant pulling force and radiative boundary conditions.

### 3. Models and methods

For the analyses presented in this chapter, an additional protein, the so-called Endoglucanase Cel5a from bacillus Agaradherans [125] (PDB-ID:1a3h) is considered in addition to a mutated version of the MBP. The latter is called MalE262 as it is mutated on residues 263 and 264. The mutation is obtained by simply quenching the 14 native contacts belonging to the issued beads. Thermalization and translocation simulations are performed for both MalE262 and 1A3H. The former ones are addressed to establish the stability of protein structures, the latter ones aim to further study the transport through the model pore described in section 2.2, which is tailored to allow single-file translocation only. The translocation simulation protocol has been already described in the previous chapter and it is adopted for both MalE262 and 1A3H transport runs. Simulations are performed at critical force  $F_c$ , *i.e.* the load intensity for which the translocation probability is 50%. This condition is believed the optimal compromise between statistical accuracy and the slow translocation speed needed to both amplify the stall duration and possibly obtain a quasi-equilibrium process. Results are not presented so much in depth as for the MBP because they do not add crucial information. However, features such as the translocation stalls will be shown along with the exposition of the 1D-continuum approach results. Such results stem from the achievement of the free energy profile in a meaningful reaction coordinate  $Q$  by umbrella sampling simulations and WHAM method [96]. Also, the free energy profile  $G(Q)$  is used in the formulation of the potential of the mean force to perform a set of one-dimensional Langevin runs in order to assess the suitability of this approach to interpret, retain and summarize the essential features of the 3D translocation dynamics, the one in (4.3).

The umbrella potential  $V_{umb}$  reads

$$V_{umb}(Q(x)) = \frac{1}{2} k_u (Q(x) - Q_0)^2. \quad (4.6)$$

The translocation process is parametrized in terms of  $Q = N_{right} - N_{left}$ , with  $N_{right}$  and  $N_{left}$  the number of residues outside the pore on its *right* and *left* side respectively (Fig. 4.2A). Translocation is thought to proceed from left to right when the pulled terminus is the C-terminus and from right to left for N-pulled translocations. In other words, the cis-side of the pore coincides with the left side in the former case and with the right side in the latter one (the trans-side accordingly). This nomenclature is used consistently throughout, with C-pulling and N-pulling experiments proceeding, for example for the MBP, from  $Q = -370$  to  $Q = 370$  and from  $Q = 370$  to  $Q = -370$ , respectively (as already stated, the MBP counts 370 residues, *i.e.*  $Q \in [-370 : 370]$ ). The associated color code used in the figures is light gray for N-pulling and dark gray for C-pulling. The collective variable  $Q$  can be expressed as a function of the  $x$ -axis coordinates of the residues in the following way:

$$Q(x) = \frac{1}{2} \sum_{i=1}^m [\tanh(ax_i) + \tanh[a(x_i - L)]], \quad (4.7)$$

with  $m$  the number of residues of the protein,  $L$  the pore length and  $a = 3.0 \text{ \AA}^{-1}$  a smoothing parameter. Equation (4.7) is a smooth version of the discrete variable  $Q = N_{right} - N_{left}$ . In order to compute the free energy profile, a set of umbrella sampling runs with different equilibrium values  $Q_0$  is performed. In particular, the interval between different  $Q_0$  is 4, hence, in the case of the MBP, 184 different runs are used to reconstruct the profile  $Q_0 = -368, -364, \dots, 364, 368$ . We find it convenient to assign different values to the elastic constant  $k_u$  of the harmonic umbrella potential depending on  $Q_0$ , in order to better control simulation outcomes. In particular, such a shrewdness allow us to explore the whole region of the phase space associated with the reaction coordinate, maintaining at the same time a good overlap between adjacent histograms and a limited amount of sampling windows. For the sake of completeness,  $k_u \in [0.22 : 2]$  for any  $Q_0$ , where the greater values are used for the higher values of  $|Q_0|$ . Indeed, higher values of  $Q_0$  correspond to the regions of the phase space where it is easier to lose the protein in the bulk even for very small fluctuations of the reaction coordinate from the equilibrium value, *i.e.* the regions in which the molecule is close to the pore mouths. Initial configurations for the umbrella sampling runs are borrowed from the configurations explored during translocation simulations by selecting the ones closer to the  $Q_0$ -value employed in the issued sampling. Each simulation runs for a time suited to collect proper uncorrelated statistics (1000 points for each histogram with a correlation time equal to 150 internal units; the latter value was determined by preliminary simulations). Moreover, the first

$10^4$  time units (equal to 10% of the translocation simulation time window) are not used for histogram collection, in order to allow the initial configuration to lose memory of the transport process and to foster the possible refolding at the trans-side. Multiple weighted histogram analysis method is applied to deweight the data arising from the procedure in order to obtain the free energy profile  $G(Q)$ . Further details on specific simulation protocols and employed quantities, such as the *contact backward density*, are given in section 4.

#### 4. Results

In this section we show that the translocation of proteins across narrow pores is generally characterized by the occurrence of long stall events, which main features are essentially retained in a 1D description of the process involving the evolution of a proper collective variable, the one in equation (4.7), in the associated free energy landscape. This 1D approach is particularly appropriate when the translocation proceeds slowly, that is when the critical driving force is low enough. Indeed, the free energy profile is just a collection of thermodynamics states along the translocation pathway, which might not be in accordance with the non-equilibrium configurations explored during the transport.

In particular it is pointed out that both the ramps along the free energy trends and the translocation bottlenecks result to be strongly associated with the structural properties of the folded proteins. In fact, they correspond systematically to the regions along the backbone chain with a dense distribution of long-range native contacts with the untranslocated portion of the protein, the so-called *backward contacts*. These contacts, as we shall see, constitute a refinement of what already introduced in section 5. In other words, the *backward contact density* along the protein structure reflects into a step-like structure of the free energy, besides determining the stalls of the transport.

The 1D Smoluchowski/Langevin approach retains the essence of the translocation phenomenon, although some unavoidable limitations, due to the simplifications introduced by the 1D modeling, can arise, especially if the *slip* transport phases in between successive stalls in the 3D molecular dynamics (MD) simulations are very fast due to a high critical force needed to perform the passage.

The overall emerging picture suggests to decompose the free energy landscape into its entropic and energetic terms. This approach confirms the importance of the regions with a dense distribution of long range interactions, which govern both internal energy and entropy trends. In fact, once such bonds are broken, we recognize not only an increment in the average value of the internal energy, but also a similar jump in the entropic term. The latter one indicates that the unfolding associated with contact severance prevails upon the entropy reduction due to the pore confinement. Such a behavior is different from the one

relevant to an unfolded protein, as the latter shows a decreasing (or stationary) entropic contribution trend.

Numerical simulations involving different proteins support the conjecture that these results are not limited to the MBP specific case, but are instead broad. For the sake of clarity, the approach is first presented referring to the MBP. Successively, the generality of the results is discussed by addressing the specific MBP mutant MalE262 and the globular protein 1A3H, which has been selected for its substantially different native topology.

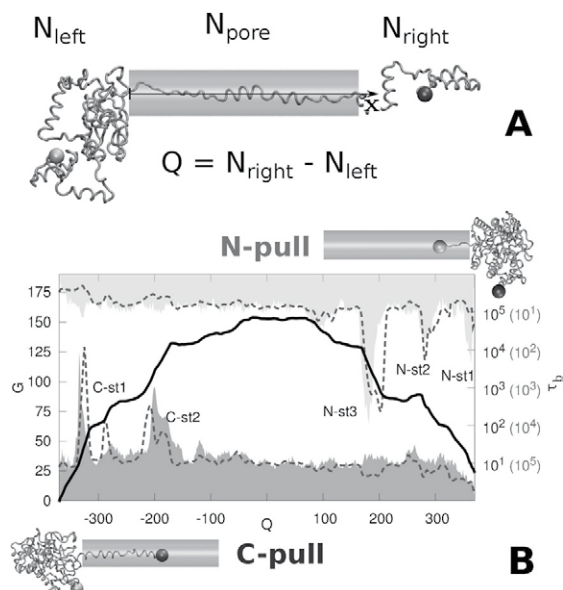


Figure 4.2: Here and in the following figures units are internal, *i.e.* related to the coarse-grained model we used. Distances, when present, are in Å. Panel A:  $N_{left}$ ,  $N_{right}$  and  $N_{pore}$  are the number of monomers on the left side, right side and inside the pore. It is always conventionally assumed that the first monomer entering the left side is the C-terminus (dark gray) while the translocation from the right side are pulled from the N-terminus (light gray). Hence the protein translocates left-to-right or right-to-left for C- and N-pulling, respectively. Panel B: Free-energy ( $G$ ) profile from umbrella sampling simulations of MBP in the collective variable  $Q = N_{right} - N_{left}$ . The curves represent the average residence time  $\tau_b(Q)$  for non-equilibrium MD simulations (filled) and for the 1D approach based on the Smoluchowski equation (4.1) (dashed). For N-pull the  $\tau_b(Q)$  scale is reversed (the values being reported in parentheses and in light gray on the right y-axis) while  $\tau_b(Q)$  stemming from the Smoluchowski equation are rescaled in order to match the background values.

The average residence time  $\tau_b(Q)$  that the MBP ( $R_c = 7.5\text{Å}$  in the present investigation) spends in the configurations associated with  $Q$  during the translocation (Fig. 4.2B), clearly indicates that the transport is not uniform. Indeed,

most time is spent in the same specific states found in chapter 3 for  $R_c = 6.8\text{\AA}$  (events C-st1 and C-st2 for C-pulling and N-st1, N-st2, N-st3 for N-pulling). In Figure 4.2B the free-energy profile  $G(Q)$  is also reported: the barriers of  $G(Q)$  correspond to the stalling events. To obtain the profile we employ the procedure described in section 3 by using as initial configurations both the conformations relative to the C and N-pulling translocation runs. In fact, from the MBP- $G(Q)$  profile obtained by using as initial configurations only those from C-pull runs and reported in dark gray in the upper left panel of Figure 4.3, it is clear that the average ascending trend in the region  $Q < 0$  does not correspond to an analogous average descending trend for  $Q > 0$ . Since outside the pore ( $Q = -370$  and  $Q = 370$ ) the thermodynamics conditions are the same for both sides of the channel, the relative free energy values should be equal. Such a prediction is not reproduced by a profile achieved by using only initial configurations which belong to either C or N-terminus translocation runs (single profile). This is presumably due to the long time needed for the refolding process at the trans side to take place. Indeed, the same behavior is confirmed for the MBP N-pull runs (light gray curve in the middle panel of Fig. 4.3 left side): for  $Q > 0$  (we conventionally consider the N-terminus pulled translocations to proceed from right to left) the ascending trend overwhelms the decreasing one observed when  $Q < 0$ . In order to obtain a single representative  $G(Q)$  profile for the entire process, half

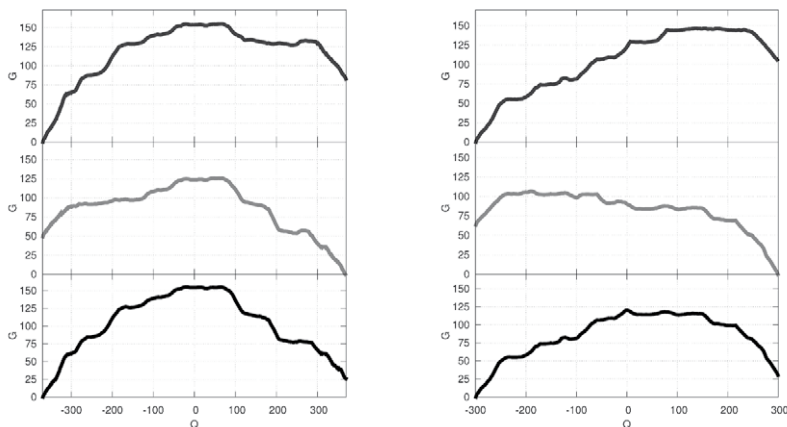


Figure 4.3: Free energy profiles  $G(Q)$  obtained from umbrella sampling runs employing initial configuration taken only from C-pull (dark gray) or N-pull (light gray) translocation runs. The black line in the lower panel is the profile obtained using the two previous profiles half by half. The left panels refer to MBP and the right ones to 1A3H.

of the umbrella sampling runs obtained from the C-terminus pulling initial conditions ( $Q_0 < 0$ ) are combined with the complementary simulations relevant to the N-terminus case ( $Q_0 > 0$ ). The resulting profile is plotted in the lower panel in Figure 4.3 and it is the one reported in Figure 4.2B. The right panels of the



former figure refer to 1A3H, for which the same approach holds. As a further indication that the anomalies of the single profiles are associated with the time needed for the refolding process, we report in Figure 4.4 the free energy profile for an unfolded protein,  $R_c = 3.0\text{\AA}$  obtained from umbrella sampling simulations exploiting just the C-pull run initial configurations. In this case  $G \simeq 0$  also for  $Q = 370$ .

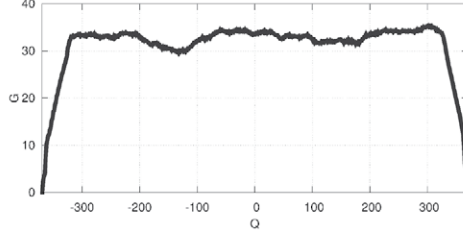


Figure 4.4:  $G(Q)$  profile for denatured MBP protein ( $R_c = 3\text{\AA}$ ) obtained from umbrella sampling employing initial configuration from C-pull translocation runs only.

As common in translocation problems, the free energy trend is used in the practical implementation of the Smoluchowski equation in order to compare the non-equilibrium MD runs with the 1D continuum approach. In this perspective, a suitable mapping of the importing force  $F_x$  employed in the MD runs on the effective force  $F_Q = \partial W / \partial Q$  used in the stochastic model is required. A gross estimate for  $F_Q$  can be obtained by assuming that, in the MD case, each position of the application point of the force along the  $x$ -axis corresponds to a unique value of the collective variable  $Q$ . In principle, from translocation experiments, it is possible to extract the conditional probability of observing  $Q$  as a function of the force application point  $x$ ,  $\hat{P}(Q|x)$ . It turns out that  $\hat{P}(Q|x)$  is sharply peaked, thus allowing to use a (monotonically increasing) deterministic function to map the MD onto the 1D-approach forces by using  $F_Q = (\partial W / \partial x)(\partial x / \partial Q) = F_x \partial x / \partial Q$ . Indeed the molecular dynamics data show that when the pore is full ( $N_{left} \neq 0$  and  $N_{right} \neq 0$ ) the distribution of  $N_{pore}$  is tight around its average value,  $\bar{N}_{pore} \simeq 38$  (Fig. 4.5) upper panels. Hence, a single monomer entering from the *cis* side corresponds to a single monomer exiting from the *trans* side (single-file motion). The net variation of  $Q$ ,  $\Delta Q = 2$ , is associated to a displacement of the force application point substantially given by  $\Delta x \simeq L / \bar{N}_{pore} \sim 2.5$ . Thus it follows that  $F_Q \sim F_x \Delta x / \Delta Q \simeq 1.25 F_x$ . Instead, when either  $N_{left} = 0$  or  $N_{right} = 0$ , we get  $\Delta Q \simeq 1$  and  $F_Q = F_x \partial x / \partial Q \simeq 2.5 F_x$ .

The dashed lines in Figure 4.2B represent the average residence time  $\tau_b(Q)$  obtained by solving a set of 1D Langevin runs governed by (4.3). It is clear that the peaks coincide with the stalls observed in the MD runs, thus confirming the accuracy of the 1D driven diffusion interpretation.

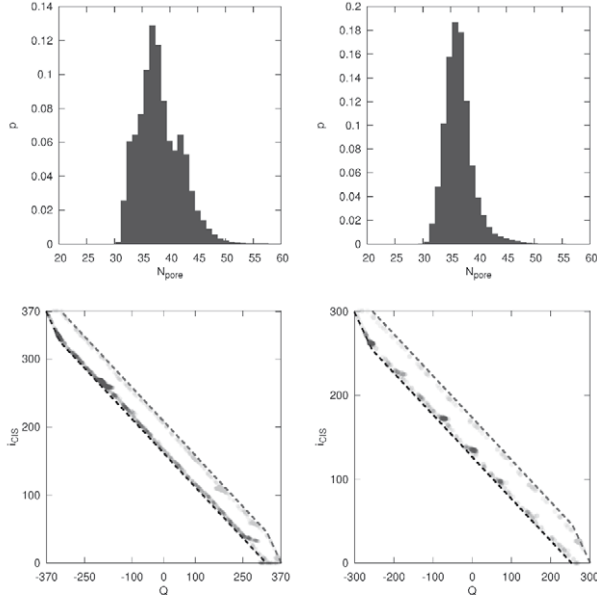


Figure 4.5: Upper left.  $N_{pore}$  distribution for translocation runs (both C and N pulling) for the MBP. Upper right, same data for 1A3H. Lower Left. Scatter plot of the amino acid at the pore mouth on the cis-side as a function of the reaction coordinate  $Q$  during MBP translocation simulations (C-pull dark gray, N-pull light gray). The maps eqs. (4.8) and (4.9) are also depicted as straight dotted lines. Lower Right. Same data for the 1A3H translocations.

It has just been shown that the Smoluchowski picture is appropriate to describe the essence of the translocation process. Now we show also that the main features of both  $G(Q)$  and  $\tau_b(Q)$  are associated with specific properties of the molecule native structure. In particular, we observe that the stalling events are related to the entrance in the pore of the regions richer in long range contacts formed with the untranslocated portion of the protein. The main idea behind this approach is that only the interactions between the parts of the molecule that still dwell on the cis-side are effective in provoking the stalls. As a first attempt to quantify such a remark, we calculate for each amino acid the *backward contact index*  $B$ , i.e. the number of contacts that the issued amino acid forms with the still untranslocated residues. First of all it is necessary to establish a map between  $Q$  and the amino acid at the pore mouth, in order to express directly the backward contact density as a function of  $Q$ , alike  $\tau_b$  and  $G$ . This view allows comparing the native contact clusters with the features of the free-energy profile, even without resorting to a dynamical analysis of the MD data. However, the average number of residues inside the pore, when it is full, is required to be, with a good approximation, a constant of the motion. In Figure 4.5 a scatter plot of  $Q$  versus the amino acid at the *cis* pore mouth,  $i_{CIS}$ , is also shown in the

lower panels. Such an amino acid is selected by running back upon the chain of the protein, starting from the foremost residue along the pore axis up to the first residue which  $x$ -coordinate falls in the range  $[-1 : 3]$ . Also, its distance from the pore axis has to be lower than the channel radius. For each  $Q$  the distribution of the amino acids that are found at the pore mouth is very tight (the map is practically one to one and onto). The dashed lines show such a correspondence by assuming  $\bar{N}_{pore} = 38$  and by using (4.8), or (4.9) to map  $Q$  and  $i_{CIS}$ . The map for the C-pulling case is

$$\begin{aligned} i_{CIS} \in [1 : \bar{N}_{pore}] & \quad Q = i_{CIS} - m, \\ i_{CIS} \in (\bar{N}_{pore} : m] & \quad Q = 2i_{CIS} - \bar{N}_{pore} - m, \end{aligned} \quad (4.8)$$

while the N-pulling map reads as

$$\begin{aligned} i_{CIS} \in [m : m - \bar{N}_{pore}] & \quad Q = -i_{CIS}, \\ i_{CIS} \in (m - \bar{N}_{pore} : 1] & \quad Q = m - \bar{N}_{pore} - 2i_{CIS}. \end{aligned} \quad (4.9)$$

Such a correspondence is mainly used to map the backward contact index relative to the amino acid at the pore mouth onto the reaction coordinate  $Q$ , by assuming that  $\bar{N}_{pore}$  is a constant of the translocation dynamics.

For a protein pulled from the N-terminus (first residue), the backward contact index for the  $i$ -th amino acid reads

$$B_i^{(N)} = \sum_{j=i+b_t}^m M_{ij}, \quad (4.10)$$

where  $M_{ij}$  is the contact matrix ( $M_{ij} = 1$  if residues  $i$  and  $j$  are in native contact and 0 elsewhere) and  $b_t$  a threshold introduced in order to consider only the native contacts formed with the residues that are relatively far in the sequence. Here we use  $b_t = 20$ , a distal distance that excludes, for instance, the interactions responsible for helix structures. Indeed, such structures do not play a considerable rôle since helices are not required to completely unfold to enter the pore. A smoothing procedure (running average) is then performed in the interval  $[i - 5 : i + 5]$ . The resulting function is indicated as  $\tilde{B}_i^{(N)}$ . Such a procedure is a refinement of what already illustrated in section 5 as, on one hand, only the backward distal contacts are considered and, on the other hand, no dynamical data are required aside from a preliminary evaluation of  $\bar{N}_{pore}$ . An analogous expression holds for C-pulling:

$$B_i^{(C)} = \sum_{j=1}^{i-b_t} M_{ij}. \quad (4.11)$$

Of course, when the amino acid  $i$  is supposed to be at the pore mouth, we get  $i = i_{CIS}$  and the maps (4.8) or (4.9) between  $i_{CIS}$  and  $Q$  hold. In Figure 4.6

the smoothed version  $\tilde{B}$  of the backward contact index  $B$  is reported for C and N-pull from just a static analysis of the MBP contact map. It is noticeable that the ascending ramps in the profile correspond with the peaks in the backward contact density  $\tilde{B}$ .

In order to confirm this picture, we remove selectively the native contacts own of the amino acids 263 and 264 (two residues mainly involved in the stall C-st2) and repeat both the non-equilibrium MD translocations and the entire umbrella sampling protocol. The result is that the residence time peak C-st2 is smaller (Fig.4.7) and, correspondingly, the  $G(Q)$  slope in that region is lowered (Fig. 4.6).

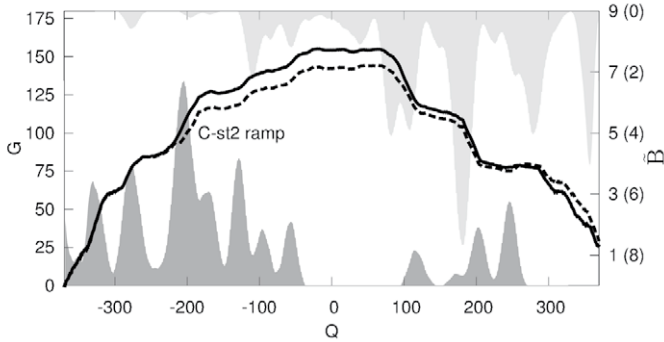


Figure 4.6: *Burial backward  $\tilde{B}$*  for C (dark gray) and N-pull (light gray). The solid line is the free energy profile of the MBP (already reported in Fig. 4.2B) while the dashed one is the profile of the MBP-mutant, for which the native contacts of amino acids 263 and 264 are removed.

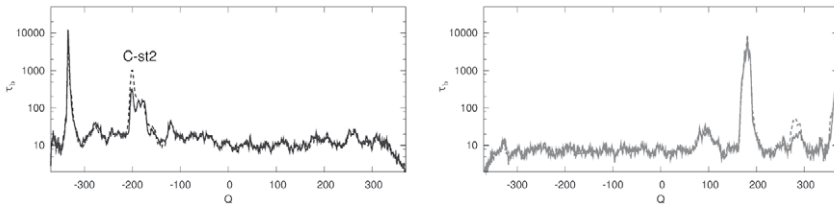


Figure 4.7: Comparison of residence time  $\tau_b$  as a function of the collective variable  $Q$  for wild-type MBP (dashed lines) and MalE262 (solid lines) for both C (dark gray) and N (light gray) pulling.

To understand if the triple correspondence between average residence time  $\tau_b(Q)$  from non-equilibrium runs, free energy profile  $G(Q)$  and protein structural features (*i.e.* the backward contact density  $\tilde{B}$ ) is specific of the MBP or is a

more widespread property of protein-like structures, we repeat the analysis for a different globular molecule (1A3H) characterized by a substantially different native topology with respect to the MBP (see Fig. 4.8 for a comparison between contact maps;  $R_c = 7.5\text{\AA}$  for the MBP and  $R_c = 7.2\text{\AA}$  for 1A3H). In particular, native contacts are more uniformly distributed for the MBP and are mainly associated with either distal clusters or  $\beta$ -sheet interactions. In contrast, for the 1A3H molecule, Lennard-Jones potentials concern primarily adjacent  $\alpha$ -helices, in addition to a small set of clusters that the region close to the N-terminus forms with the rest of the structure. In the bottom panel of Figure 4.8 the quenched contacts in MalE262 are highlighted in dark gray. The protein 1A3H is formed

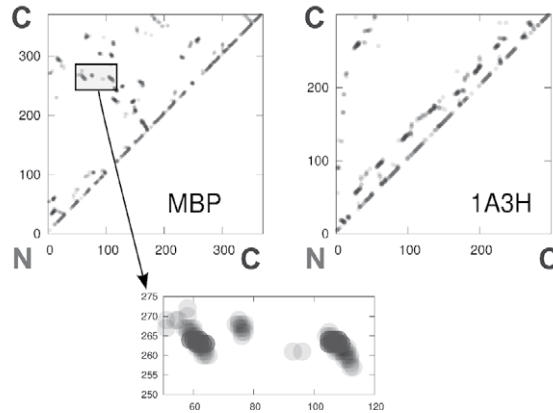


Figure 4.8: Native contacts are reported in transparent light gray circles. The dark gray circles in the bottom panel correspond to the fourteen contacts removed in order to obtain the mutated Maltose Binding Protein, MalE262.

by 300 residues (hence  $Q \in [-300, 300]$ ). In Figure 4.9,  $G$ ,  $\tau_b$  and  $\tilde{B}$  for both C (dark gray) and N-pull (light gray) are reported as a function of  $Q$ . The dynamics, similarly to the MBP, is characterized by several stalls. For both N-pull and C-pull it appears that the peaks in  $\tau_b(Q)$  coincide with the ones in the corresponding  $\tilde{B}$  trend. Moreover, the larger clusters of native contacts are associated with the barriers in  $G(Q)$ .

Here the Smoluchowski picture (dashed lines) is appropriate to summarize the 3D phenomenology, mainly for the C-pulling case, as the N-pull is characterized by a very high value of the critical force. Such a force probably reduces the suitability of the 1D approach in the equilibrium potential dictated by the free energy profile, at least from a quantitative point of view. In Table 4.1 critical forces  $F_c$  are summarized for the translocation simulations performed. For the mutated MBP the critical force can be assumed to be equal to the one pertinent to the wild type structure.

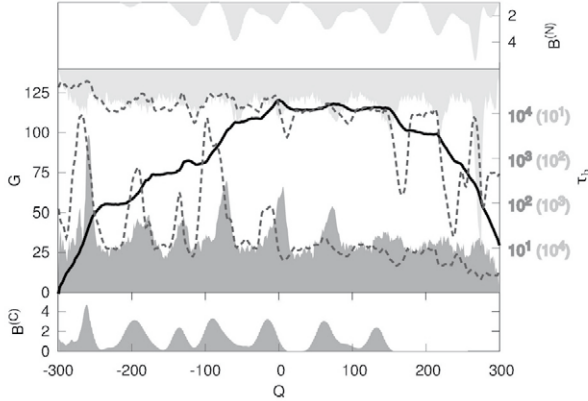


Figure 4.9: 1A3H. The filled curves in the main panel represent the average residence time  $\tau_b$  from 3D MD translocation runs, while the dashed lines represent the outcome of the 1D Smoluchowski approach. The solid line is the free energy profile  $G(Q)$  from umbrella sampling simulations. The upper and lower panels report the *backward contact index*  $\bar{B}$  for C (dark gray) and N-pull (light gray).

Protein, Pull	$F_c$
MBP C-pull	2.1
MBP N-pull	2.9
1A3H C-pull	1.9
1A3H N-pull	3.6

Table 4.1: Critical forces for the different MD translocation runs performed. The critical force of the mutated MBP coincides, within the statistical error, with the critical force of the corresponding wild type structure.

Finally, we analyze the different contributions to  $G(Q)$  in order to understand what are the dominant ones for the observed phenomenology. In Figure 4.10A, the total internal energy  $V_{G\bar{o}}$ , the free energy  $G$  and the entropic contribution  $TS$  are reported for the wild type MBP as a function of  $Q$ . The energetic contributions are calculated as conditional average by using the configurations from the umbrella sampling runs. The entropic term is  $TS(Q) = V_{G\bar{o}}(Q) - G(Q)$ . All contributions are conventionally set to zero for  $Q = -370$  (protein on the left side of the pore). During the transport  $V_{G\bar{o}}$  increases since the protein needs to unfold to translocate and, consequently, native contacts are stressed up to rupture. A similar behavior is found for the entropy. Indeed the break-up of native contacts results in a less compact structure and in a disorder increment, up to a maximum corresponding to the protein straddling the pore symmetrically (the unfolding associated with contact detachment overwhelms the confinement effect of the channel). Entropic and energetic contributions are

closely correlated one another and, as usual, with the clusters of native contacts. Also, the entropy profile differs from the one pertinent to an unstructured polymer, which is provided in Figure 4.11 upper panel (MBP  $R_c = 3.0\text{\AA}$ ). As it is expected, the entropic term shows just a barrier at the beginning of the translocation, *i.e.* up to the exit of the first residue on the *trans* side, after which the landscape is practically flat. Panel B of Figure 4.10 reports the main contributions to  $V_{G\bar{o}}$ , namely  $V_{nat}$  and  $V_\phi$ . The dashed line  $V_{nat}^*$  marks a heuristic estimation for  $V_{nat}$ , obtained by integrating the balanced version of the backward contact density subsequently described.

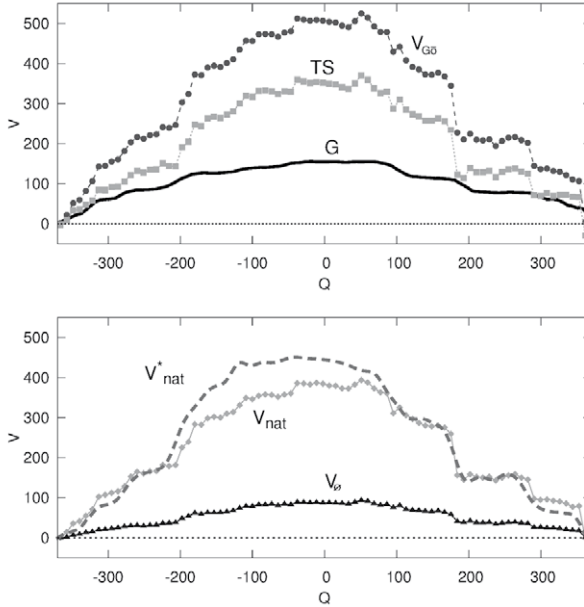


Figure 4.10: Upper: Energetic (dark gray circles) and entropic (light gray squares) contributions to the free energy profile  $G(Q)$  for wild type MBP. Lower: Main contributions to  $V_{G\bar{o}}$ :  $V_{nat}$  (light gray diamonds) and  $V_\phi$  (dark gray triangles). The dashed line represent the heuristic estimation of  $V_{nat}^*$  based on the backward contact density.

For the sake of clarity, a translocation pulled from the C-terminus is exemplified. A quantitatively identical result can be obtained by considering the N-pull translocation. As first step we calculate the quantity

$$\tilde{B}^{(C)*}(Q) = \tilde{B}^{(C)}(Q(i)) - \tilde{B}^{(N)}(Q(i + \bar{N}_{pore})), \quad (4.12)$$

where the term  $\tilde{B}^{(N)}(Q(i + \bar{N}_{pore}))$  represents the contribution to the internal energy due to the refolding on the *trans* side. In fact, when residue  $i$  dwells at the pore mouth on the *cis*-side, residue  $i + \bar{N}_{pore}$  dwells close by the *trans*-side

pore entrance. Integration of (4.12) provides an estimation for the net number of broken contacts during the translocation. Since each contact amounts for an energy jump equal to 1, a rough estimation of the native contribution to the internal energy can be calculated as

$$V_{nat}^*(Q) = \int_{-m}^Q \tilde{B}^{(C)*}(q) dq. \quad (4.13)$$

The approach is a further confirmation that the essence of the translocation is hidden in the structural properties of the native configurations, which are in turn mainly determined by the long range native interactions.

To conclude, we mention that analogous results in terms of energetic and entropic contributions hold for the 1A3H protein (Fig. 4.11 lower panel), thus confirming the overall picture.

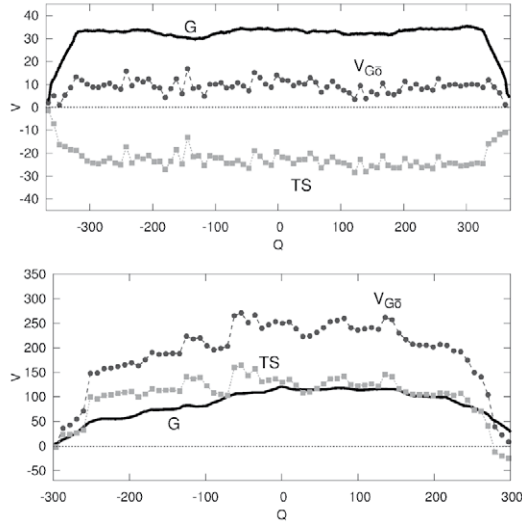


Figure 4.11: Energetic (dark gray circles) and entropic (light gray squares) contributions to the free energy profile  $G(Q)$  for unfolded MBP ( $R_c = 3 \text{ \AA}$  upper panel) and for 1A3H (lower panel).

## 5. Remarks

In this chapter it has been shown that the stalling events that characterize the dynamics of a protein-like structure in a single-file translocation are associated to the structural properties of the native configuration. The general setting holds for all the cases analyzed. However, the stall sequence is specific for each



protein and constitutes a sort of signature of the molecule. The identification of the features responsible for the bottlenecks of the transport allow us to formulate a heuristic procedure able to estimate the native bound contribution to the free energy profile, on the sole basis of the protein native structure and  $\bar{N}_{pore}$ , by simply integrating a balanced version of the backward contact index profile. These occurrences appear to be generic, not associated just to a specific globular protein, thus opening the way to systematic pre-screening of the proteome.

Also, the intrinsic difference between folded and unfolded structures in terms of the entropic contribution relevant to the translocation process has been pointed out. In the former case the confinement effect due the pore appears to be overcome by the severance of native contacts, hence confirming the importance and leading role of the long range interactions in this kind of dynamics. On the other hand, for unstructured polymers, the free energy landscape is dominated by a single entropic barrier at the beginning of the transport.

Finally, the one-dimensional interpretation of the process as a driven diffusion along a proper potential of the mean force provides overall good results. Indeed, it retains the essence of the 3D translocation dynamics like, for example, the bottleneck scenario. However, quantitative comparisons appear to be too ambitious, especially if the critical force pertinent to the transport is high, as for the 1A3H N-terminus pulling. In fact, in this case, once the first stall has been overcome, the translocation proceeds extremely fast and hence it is poorly described by a quasi-static approach.

# Chapter 5

## 3-Dimensional Continuum Models

### 1. Preamble

In the present chapter we propose a continuum model for the description of the dynamics of isolated macromolecules. We adopt a second-rank tensor as a descriptor of the macromolecular shape and identify the action governing its dynamics by means of an identification procedure from a discrete scheme, based on power equivalence and Cauchy-Born rule. We compare molecular dynamics stretching simulations with the continuum model by starting from discrete toy schemes, going on with increasing complexity, and ending with the analysis of the already encountered Ubiquitin protein. The results indicate limitations in the approach in case of unconstrained molecular dynamics while they show appropriateness for driven dynamics.

In the second part of the chapter, simulations relevant to the MBP protein are also presented. In this case stretching and, above all, translocation dynamics are considered. The selected approach is able to capture the shape evolution of the molecule, providing that the pore is long enough to provoke a considerable stretching. However, limitations totally analogous to the Ubiquitin protein (and hence not shown for brevity) still hold for equilibration simulations.

Such shortcomings are overcome thanks to empirical formulations of the morphological descriptor, or rather of the quantity that governs its dynamics, as we will see.

### 2. Introduction

Biological materials have in general intricate architecture at finer spatial scales. Multi-scale and multi-field views are often called upon to offer a setting for the description of their mechanical behavior. Even the analysis of processes involving a single macromolecule requires care. This is the case, for example, of protein stretching in free space (performed, for example, in atomic force microscope experiments). In analyzing such a process, both discrete and continuous views are adopted. Numerical simulations on discrete schemes (both full atoms and coarse-grained molecular dynamics) are developed to analyze stretching and transport (see, e.g., [54, 55, 123, 126, 127]), and to

determine free-energy profiles (see, e.g., [11, 55, 128, 129]) which are used often to obtain closed form solutions for first passage time statistics (see [57] for relevant theoretical issues).

Models at continuum level are then used to summarize the results of numerical simulations, by reducing the dynamics of discrete material points to the one of a single descriptor of the molecule. The common view is that of using data to describe just the motion of the molecular mass center or of a similar collective variable leading to one-dimensional dynamics (see, e.g., [11, 55], and references therein). No information on molecular shape evolution in time are added.

Here we propose a continuum model which captures, at least in a coarse way, the shape evolution of an isolated protein. So, we attribute to the molecule a (time-dependent) second-rank tensor  $\nu$  with positive determinant as a descriptor of its shape.  $\nu$  does not coincide with the inertia tensor. It is a linear operator that we define on the basis of Cauchy-Born rule. In fact, we imagine to fix a shape of the molecule that we take as a reference, and represent it by a discrete set of material points.  $\nu$  is then defined as the tensor such that the velocity of the  $i$ -th material point is given by the time rate of  $\nu$  applied to the  $i$ -th reference position vector.

The evolution of  $\nu$  is governed by a balance of (microscopic) actions on the molecule. At continuum level we identify such actions in terms of the ones occurring between every material point and its first, second, and third neighbors in the discrete scheme, in addition to possible long-range interactions. The identification is made in terms of power equivalence. The values of the actions in the discrete model are obtained by numerical simulations. Then the continuum evolution of  $\nu$  is depicted in terms of the eigenvalues of the symmetric component of its polar decomposition.

We compare molecular dynamics stretching simulations with the continuum model. We consider first discrete toy schemes and go on with increasing complexity, then we end with the analysis of the Ubiquitin protein. The results indicate limitations in the approach in case of free dynamics of the macromolecule while they show appropriateness for constrained dynamics, allowing us to accept the model for a number of constrained motions such as protein mechanical unfolding.

Finally, MBP translocation is also taken into account (there is not substantial difference between Ubiquitin and MBP thermalization and mechanical stretching results, hence MBP data are not reported for brevity, aside from just one remark in sec. 4.4). We shown that the appropriateness of the approach still holds in MBP driven and confined dynamics. At the end of the chapter, we also point out that the limitations concerning free equilibration are overcome by selecting empirical formulations of the so-called *self-action* (a quantity that, in the present setting, governs the evolution of  $\nu$ ).

### 3. Models and methods

#### 3.1. Discrete scheme

The Ubiquitin protein in Figure 5.1 is described by the already introduced (sec. 2.1 of chapter 3)  $C_\alpha$  backbone coarse-grained Gō-like model (see [1, 100] for the original approach, [2] for the refined version implemented here and also [2, 55, 56, 114], for its use in describing protein stretching, thermalization and translocation). The scheme is also equivalent to what is presented in [55] where the same molecule (PDB-id:lubq) is considered. Values of the parameters appearing in protein potentials and measure scales are reported also in [55]. If not

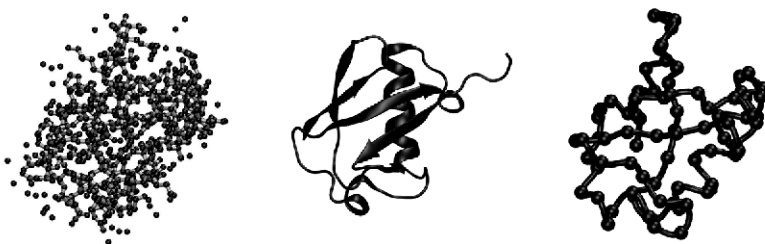


Figure 5.1: From left to right: full-atom picture, secondary structure features and coarse-grained  $C_\alpha$  carbon atoms along the backbone chain.

differently specified, we assume also  $R_c = 3.0\text{\AA}$  or  $R_c = 6.5\text{\AA}$  as values of the Gō model cut-off radius. The Gō potentials (see sec. 2.1 of chapter 3) constitute a basis for modeling atomic interactions not only in the Ubiquitin structure but also in the other sets of atoms in space, as it will be pointed out in what follows when necessary. The same approach holds also for the MBP.

#### 3.2. Continuum model

As anticipated in the introduction, at continuum level the dynamics of a single protein is commonly depicted just by evaluating the subsequent positions in time of its centre of mass (see e.g. [11]). A more refined natural descriptor of the molecular morphology is the moment of inertia of the atomic cluster. When we use it as descriptor of the molecule, as Ericksen suggested in [58], its time evolution involves the sum of the tensor moment of momentum of each single atom constituting it.

We do not follow such a view, but we are inspired by it and select a second-rank tensor as a descriptor of the molecular shape, a tensor associated with an idea of homogeneous deformation of a region containing the molecule in a shape

that we take as a reference (in this sense we are closer to what suggested in [59] for the mechanics of granular media). So, we select a compact region  $\epsilon$  in three-dimensional space  $\mathbf{R}^3$  – the volume of  $\epsilon$  is indicated by  $|\epsilon|$  – as a large enough portion of space containing the molecule in a shape that we take as reference (the equivalent sphere radius ranges in the interval  $[15,25]\text{\AA}$ , about one and a half to two times greater than the Ubiquitin gyration radius). Let  $r_i$  be the vector position of the  $i$ -th atom – or material point if we refer to a coarse discrete representation of the protein – in  $\epsilon$  in the reference place. We consider a second-rank tensor  $\nu$  with positive determinant, which depends on time  $t$ , such that at the instant  $t$  the velocity  $v_i$  of the  $i$ -th material point in the discrete scheme is simply given by  $\dot{\nu}r_i$  – essentially we are imagining that  $\epsilon$  is endowed with an homogeneous strain rate and the atoms of the molecules are dragged along the process by the homogeneous deformation.

The assumption has to be interpreted ‘in average’. It has clear limitations because the unfolding of a protein can develop in a non-homogeneous way by involving subsequent portions. However, the assumption

$$v_i = \dot{\nu}r_i, \quad (5.1)$$

which is motivated by Cauchy-Born rule<sup>1</sup>, allows us to identify clearly the action of the protein over itself.

We presume also that  $\nu$  be endowed with positive determinant, so that, by polar decomposition theorem, we can write

$$\nu = RU, \quad (5.2)$$

where  $R \in SO(3)$  and  $U$  is a second-rank symmetric tensor.

In the discrete representation of the macromolecule, as described previously, call  $f_{ij}$  the action (a force, a covector indeed, that we think derived from a potential, the one introduced previously) between the material points  $i$  and  $j$ , applied to the point  $i$ . The density of power  $\pi_{discr}$  developed in the discrete scheme is then given by

$$\pi_{discr} := \frac{1}{|\epsilon|} \sum_{i \neq j \in \epsilon} f_{ij} \cdot v_i.$$

In the continuum representation, the rate  $\dot{\nu}$  has a conjugated action which is defined by the power  $\pi_c$  which is necessary to develop over the molecule, considered as a whole, to get the rate  $\dot{\nu}$  itself. Since  $\dot{\nu}$  is a tensor, the power-conjugated action is a tensor too – call it  $\bar{z}_s$  – so that  $\pi_c$  is defined by

$$\pi_c := \bar{z}_s \cdot \dot{\nu}.$$

<sup>1</sup>See [130] and [131] for extended analyses of such a rule.

Moreover, since  $\bar{z}_s$  summarizes the actions in the protein (it is a self-action, indeed), and we are not considering for the moment other actions over it, interaction with the external environment being assigned, in this case, by applied strain to the ends of the protein,  $z_s$  should satisfy the balance equation

$$\bar{z}_s = 0. \quad (5.3)$$

However, in principle  $\bar{z}_s$  could be imagined as the sum of a conservative part,  $z_s$ , and a dissipative one,  $z_s^d$ , dissipation being developed during the evolution of the molecule – so we should have  $\bar{z}_s = z_s + z_s^d$ . The dissipative nature of  $z_s^d$  would be then expressed by the inequality

$$z_s^d \cdot \dot{\nu} \geq 0,$$

which should hold for any choice of  $\dot{\nu}$ , implying so an expression for  $z_s^d$  of the type

$$z_s^d = a\dot{\nu}^b, \quad (5.4)$$

with  $a$  a positive constant and  $\dot{\nu}^b$  the dual version of  $\dot{\nu}$ . Precisely, with  $e_1, e_2, e_3$  a vector basis in the three-dimensional ambient space, and  $e^1, e^2, e^3$  its dual counterpart, we have  $\dot{\nu} = \dot{\nu}^{ij} e_i \otimes e_j$  and  $\dot{\nu}^b = \dot{\nu}_{ij}^b e^i \otimes e^j$ . The reason of the occurrence of  $\dot{\nu}^b$  in (5.4) is that this way the product  $a\dot{\nu}^b \cdot \dot{\nu} = a\dot{\nu}_{ij}^b \dot{\nu}^{ij}$  is naturally defined without the additional structure of the scalar product. However, in what follows we shall not make difference between covariant and contravariant components because we shall always tacitly use the identification of  $\mathbf{R}^3$  with its dual. As a consequence the balance (5.3) would reduce to the evolution equation

$$a\dot{\nu} = -z_s.$$

In particular, for the sake of simplicity we shall assume later that  $a = 1$ , so that the evolution equation that we shall consider will be

$$\dot{\nu} = -z_s. \quad (5.5)$$

If so, the power  $\pi_c$  would result the sum

$$\pi_c = z_s \cdot \dot{\nu} + \frac{1}{2} |\dot{\nu}|^2.$$

The problem is now the explicit expression of  $z_s$  in terms of the discrete actions  $f_{ij}$ . To get it, we impose here the identification of  $\pi_{discr}$  with  $z_s \cdot \dot{\nu}$ : the power of the (potential dependent) actions in the discrete scheme should be the one of the conservative part of the self-action. We then write

$$z_s \cdot \dot{\nu} = \frac{1}{|\mathbf{e}|} \sum_{i \neq j \in \mathbf{e}} f_{ij} \cdot \mathbf{v}_i.$$

Then, thanks to the arbitrariness of  $\nu$ , by using (5.1) we finally get

$$z_s = \frac{1}{|\mathbf{e}|} \sum_{i \neq j \in \mathbf{e}} f_{ij} \otimes r_i. \quad (5.6)$$

In presence of (conservative) external forces on single material points of the discrete representation of the molecule, the previous expression of  $z_s$  should be augmented by the tensor product of these forces by the position vectors of the material points where they are applied. So, if  $\bar{f}_k$  is the (conservative) force applied to the  $k$ -th material point, with  $k = 1, 2, \dots, q$ , with  $q$  the number of atoms where external action act, we then have<sup>2</sup>

$$z_s = \frac{1}{|\mathbf{e}|} \left( \sum_{i \neq j \in \mathbf{e}} f_{ij} \otimes r_i + \sum_{k=1}^q \bar{f}_k \otimes r_k \right).$$

Since we are interested in the changes in molecular shape, in the numerical analyses presented later we shall compute essentially the eigenvalues  $\lambda$  of the symmetric tensor  $U$  appearing in the polar decomposition of  $\nu$ , and determining the semi-axes of an ellipsoid in three-dimensional space.

Finally, in what follows, the placement vectors  $r_i$  might be referred to different frames. However, in the case of Ubiquitin and Maltose Binding proteins, only placement vectors with respect to the center of mass of the molecules will be considered. Obviously, explicit coordinates will depend on the frame selected to evaluate both the center of mass and bead positions.

### 3.3. Stretching simulations: methods

We summarize here the methods that we use in developing later numerical simulations.

Stretching simulations are performed by means of a scheme based on constant velocity steered molecular dynamics (see [117]). Initial conditions are obtained by the so-called ‘free equilibrium’. At the beginning of each numerical test the tensor  $\nu$  is set to be equal to the identity. Protein detachment is achieved by a spring of elastic constant  $k = 1\epsilon$ , linking a fictitious atom with a terminus of the molecule. The fictitious atom is pulled at constant velocity. Its motion drags the terminus where it is bonded. We perform two different stretching protocols.

1. In the first scheme one end is held fixed in space and only the other one is pulled away.
2. In the second approach both ends of the molecule are dragged with opposite velocities.

<sup>2</sup>For an enlarged view on this identification procedure in the setting of mechanics of complex materials, see [132].

The stretching direction coincides with the vector joining the two ends of the molecule in the initial configuration. We investigate different stretching velocities  $v_{str}$ , temperatures  $T$  and cut-off radii  $R_c$ . When maximum elongation is reached (the molecule is in an almost single-file conformation), the protein can be either kept stretched or left free to fluctuate.

For the other smaller sets of atoms, either the fictitious atom external spring (the one connecting the fictitious material point to the molecule) or the direct enforcement of a suited constant velocity motion to some points of the system have been used as strategies to obtain a deformation dynamics.

### 3.4. Translocation simulations: methods

As far as MBP translocation simulations are concerned, nothing changes from what described in sec. 3.1 but the pore length, which ranges in the interval  $[100 : 800] \text{ \AA}$ , depending on the simulation performed.

## 4. Results

### 4.1. Toy simulations and related remarks

Before looking at the dynamics of the entire Ubiquitin molecule at discrete scale and within the continuum approximation proposed, we analyze some simpler schemes – toys, indeed – to investigate the limits of the discrete-to-continuum identification that we have adopted here.

**1D biatomic systems.** First we consider the evolution of  $\nu$  for a trivial harmonic oscillator, composed by two material points, indicated by  $A$  and  $B$ , and connected by a spring of elastic constant  $k_h$ . In this case  $\nu$ , defined previously, reduces to a scalar. The scheme, sketched in panel *a* of Figure 5.2, allows us to make some elementary predictions that are in agreement with the numerical analyses. The simulation temperature is set equal to zero and constant velocities are assigned to the two atoms.

Let  $r_0$  be half of the inter-atomic distance when the spring between points  $A$  and  $B$  is at rest (in other words, in a configuration that we set as a reference), that is  $r_A(t = 0) = -r_0$  and the same is for  $B$ .

When two equal and opposite velocities  $\pm v_0(t)$  are assigned to points  $A$  and  $B$ , for the motion we get

$$r_A(t) = -r_0 - \int_{t_0}^t v_0(\tau)(\tau - t_0)d\tau, \quad r_B(t) = -r_A(t).$$



## Coarse-grained molecular dynamics and continuum models

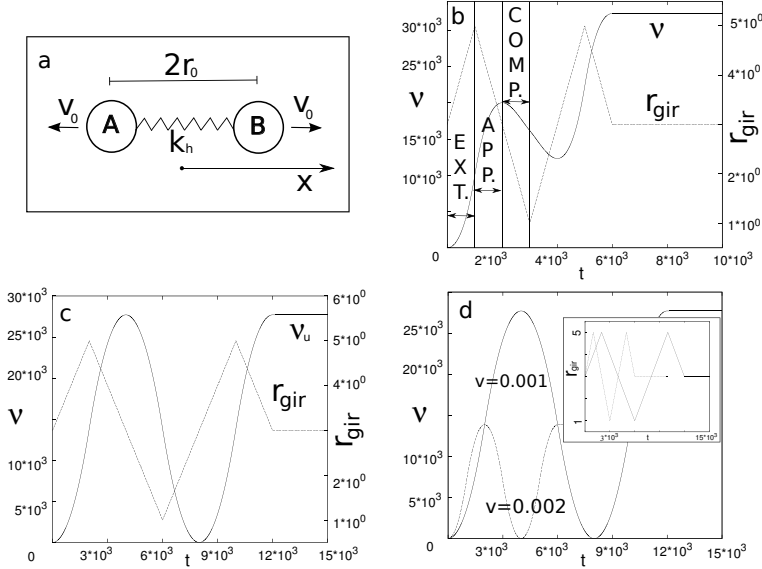


Figure 5.2: Two-atoms system. Here and in all the other figures lengths are expressed in  $\text{\AA}$  and the time in the relevant 'code unit'. Panel *a*: Sketch of the two-atoms mass-spring system. Panel *b*: Evolution of  $\nu$ , a scalar in this case, for a standard elongation/compression cycle, run twice. The asymmetric evolution of  $\nu$  would generate an almost-monotonic trend if the cycle was performed over and over again. The evolution of  $\nu$  is here compared with the radius of gyration  $r_{gir}$  of the set of atoms. Panel *c*: It is the same of panel *b* except that here the specular evolution of  $\nu$  is achieved because internal distances  $r_i^G$  appearing in the expression of  $z_s$  are kept constant to the reference values during the whole dynamics. We note a delay in the evolution of  $\nu$  with respect to the trend of the radius of gyration as the former keeps on increasing (or decreasing) during the approaching phases (*i.e.* when the system is moving toward the rest position from maximum detachment or compression). Also,  $\nu$  never shrinks further than its initial state. Panel *d*: Effect of the velocity on the evolution of  $\nu$ . Despite an overall equivalent deformation testified by the evolution of the gyration radii in the inset, the amplitude of the trends of the ellipsoid is different if the velocity is changed, *i.e.* it is dependent on the rate of deformation.

By indicating with  $f_{AB}$  the force that point *A* exerts on point *B*, we can write

$$f_{AB} = -k_h(r_B(t) - r_A(t) - 2r_0), \quad f_{BA} = -f_{AB}.$$

We develop an elongation cycle in which points *A* and *B* are pulled far apart by imposing opposite constant velocities  $\pm v_0$  to *A* and *B*. The initial position with the atoms at rest is obviously regained by reversing the velocities for the same amount of time spent in the elongation phase, *i.e.*  $t_e = t_a$ , where the index *e* indicates the elongation time and *a* the time spent in the re-approaching process. The analysis of the previous equations is trivial and the solution can be plugged into the definition of  $z_s$  (where for simplicity we take  $|\epsilon| = 1$ ).

Assume  $t_0 = 0$ . From the coarse grained balance of overall actions in the continuum scheme, we find two expressions for  $z_s$  that we label with the apex  $e$  in the elongation phase, and with  $a$  when the two particles re-approach one another by reverse velocity. We get

$$z_s^e = -4k_h(r_0v_0t + v_0^2t^2), \quad z_s^a = -2k_h(v_0t - 2r_0)^2.$$

Time integration of (5.5) along an extension-compression cycle starting from an initial value  $v_0$  then leads to

$$\begin{aligned} \nu_e &= \nu_0 + \int_0^{t_e} 4k_h v_0 (r_0\tau + v_0\tau^2) d\tau \\ &= \nu_0 + 4k_h(v_0r_0\frac{t_e^2}{2} + \frac{1}{3}v_0^2t_e^3) = \frac{10}{3}k_h t_e r_0^2 \\ \nu_a &= \nu_e + \int_{t_e}^{t_a} 2k_h(v_0\tau - 2r_0)^2 d\tau > \nu_e. \end{aligned}$$

The last inequality arises for  $z_s^a$  is negative in the whole approaching process. So, we have a drawback: the evolution of  $\nu$  is monotonic although in the approaching phase the two atoms are getting closer toward the initial configuration.

Also, if we we start compressing, the effect on  $\nu$  is in magnitude different from the extension dynamics. Indeed the expression of  $z_s$  for such a case is

$$z_s = 2k_h r_0 v_0 t - 2k v_0^2 t$$

The two terms on the right hand side are opposite in sign: the evolution of  $\nu$  results smoother and slower than in the elongation phase. In other words, a symmetric deformation process does not generate a symmetric evolution of  $\nu$ . First  $\nu$  does not recover the initial condition once the two atoms are back to the rest position. It also does not regain the starting value even when a compression follows the initial elongation, as clearly depicted in panel *b* in Figure 5.2. In other words,  $\nu$  never ‘shrinks’ below the initial value. The evolution process quickly degenerates in an almost monotonic trend, if the cycle elongation-contraction is repeated several times. Moreover, if opposite sinusoidal motions are imposed to points *A* and *B*, in place of the constant velocity drifts, a complete monotonic evolution of  $\nu$  arises, as it results from the subsequent elementary analysis.

In the sinusoidal case, the harmonic displacements (constants do not have here a prominent rôle) are given by

$$r_B(t) = r_0 + D \sin(D'(t - t_0)), \quad r_A(t) = -r_B(t)$$

The forces exerted between the points are then

$$f_{AB}(t) = -k_h(r_B(t) - r_A(t) - 2r_0) = -2k_h r_B(t)$$

and

$$f_{BA}(t) = -f_{AB}(t).$$

$z_s$  is given by

$$z_s(t) = f_{AB}(t)r_B(t) + f_{BA}(t)r_A(t) = 2f_{AB}(t)r_B(t) \propto -\sin^2(t).$$

By integrating in time (5.5), we find that  $\nu$  diverges, namely

$$\nu(t) \propto \int_0^t \sin^2(\tau) d\tau \propto t,$$

thus we have again a bounded oscillation for the material points on one hand, and an unbounded evolution on the other hand, if  $t \rightarrow \infty$ .

An attempt to eliminate this inadequacy can be the decision to fix  $r_A(t) = -r_0$  and  $r_B(t) = r_0$ , an internal constraint, indeed. In this case, in the simple 1D harmonic case that we treat here, the evolution of  $\nu$  would result just governed by the forces determined by the internal spring, and it would be bounded and symmetric. But there would still exist an half-a-period delay between the actual overall shape of the system and the associated trend of  $\nu$  at the continuum level. Namely, the evolution would be still monotonic during the elongation and contraction phases, obviously for the sign of the internal forces does not change (see panel *c* in Fig. 5.2).

From now on, all results will refer to the just described approach, namely the choice  $r_i(t) = r_i(0)$ ,  $i = 1, \dots, m$ , with  $m$  the number of material points in the cluster, if not differently stated. However, in section 4.3 we provide also an example of monotonic evolution of  $\nu$  for the Ubiquitin protein in the case this assumption does not hold, *i.e.*  $r_i = r_i(t)$ .

An additional shortcoming suggested by this simple 1D analysis arises directly from the combination of equations (5.5) and (5.6). If we suppose that the same evolution is performed in different times (e.g. we perform a whole extension-contraction cycle once at velocity  $v_0$  and once at velocity  $v_0/2$ ), although the same gross deformation occurs, different values of  $\nu$  are detected. In fact, in the slower case the same trend of  $z_s$  spreads over a time that is twice the time of the fastest one. Hence the integral results greater (doubled if the self-action evolution is linear), as pointed out in panel *d* of Figure 5.2.

Finally, it is obvious that if the biatomic 1D system considered in this section is kept elongated or compressed, its shape does not vary but, since the spring is loaded,  $z_s$  is different from zero and  $\nu$  evolves independently (data are not reported here).

If the interatomic potential among material points is changed to, for example, a Lennard-Jones one, as the one in (3.4), we do not find an improvement in the picture. Actually the very stiff Lennard-Jones potential makes it almost impossible to achieve compression without huge spikes in the evolution of  $\nu$  (see panel *a* in Fig. 5.3). Therefore, to perform such an analysis, two extra-particles

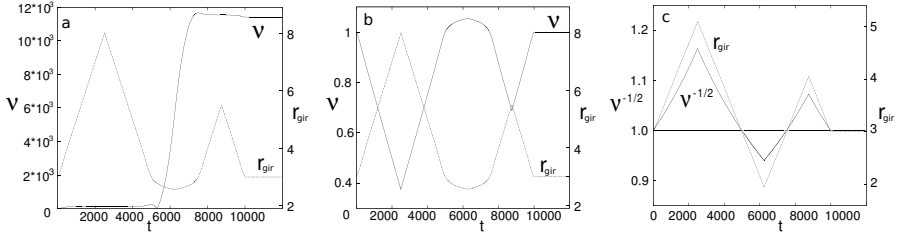


Figure 5.3: Two-atoms system. Panel *a*: Trend of  $\nu$  when the interaction between neighboring atoms is governed by a Lennard-Jones potential. The latter one is very stiff and the associated  $\nu$  undergoes a very large peak in correspondence with the maximum compression. The decreasing trend is practically negligible. Panel *b*: Evolution of  $\nu$  when external forces, due to the interaction between the system and the fictitious atoms that lead the deformation process, is considered in the computation of the self-action. Panel *c*: Dynamics of  $\nu$  when, as above, the external forces are taken into account and the square-root of the reciprocal of the square root of the order parameter  $\nu^{-1/2}$  is the quantity chosen to describe the shape evolution of the system. Here the inner interaction potential is an harmonic covalent bond, but a similar outcome is detected also in case a stiff Lennard-Jones potential is enforced. Qualitative and quantitative agreements between  $\nu$  and the radius of gyration of the 1D system are achieved.

are bonded through an harmonic potential to the atoms constituting the system. These fictitious atoms are driven at constant velocity instead of directly drifting the two inner points. In this approach we can either include or neglect the contribution of the external springs on the forces that act upon the inner atoms of the systems. Such a choice appears crucial as the external harmonic potentials end up governing the dynamics, and the related forces combined with the internal ones do practically change sign as soon as the motion is reversed (see panel *b* in Fig. 5.3). This is confirmed not only for Lennard-Jones potentials but also when the inner forces are in turn harmonic potentials (see panel *c*, same figure). In this analysis (differently from panel *b*) we take the square-root of the reciprocal of  $\nu$ , in analogy with the eigenvalues of the inertia matrix and the relevant inertia ellipsoid. This shrewdness allow us to achieve a trend in compliance with the radius of gyration and to avoid the amplitude mismatch observed in panel *b*. The order of magnitude of the quantities in comparison is also similar.

In this 1D example there is no influence of temperature on the scheme we are using, providing that the effect of the external springs are not accounted for in the force vectors determining  $z_s$ . Once the latter is included to achieve a clean representation of the numerical results, as the one in the just mentioned panel *c*, the toy scheme seems much more sensible to thermal fluctuations and to their ability to spoil the picture, due to the origin of non-homogeneous deformation terms. However low-enough values of the temperature do not remarkably alter stretching results.

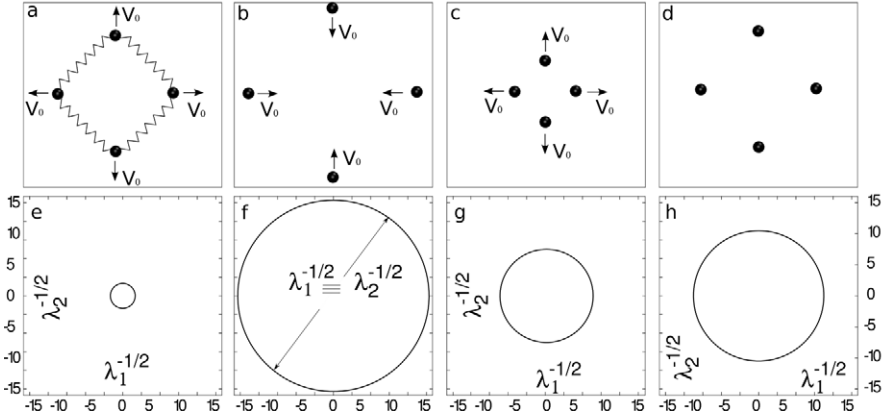


Figure 5.4: First row: Dynamics of a homogeneous deformation for a four-atoms system. Each material point is pulled at constant velocity in an extension/compression cycle. Second row of pictures: Evolution of the eigenvalues  $\lambda_i$  of the symmetric part of  $\nu$  obtained by polar decomposition. The order parameter for the shape of the system never shrinks below the initial state despite the set of atoms undergoes a compression phase (third column from the left). This shortcoming is common to all the situations where external forces are absent (this is due to the direct assignment of constant velocity motions to the points of the system) or not considered in the computation of the self-action. Obviously, since the deformation process is symmetric,  $\nu$  is a circle.

**2D four-atom system.** Here we briefly mention the outcome of 2D simulations performed over a set of four atoms in space, placed at the corners of a square. Once homogeneous deformation is applied (all atoms are suitably driven by constant velocity motions) the overall basic picture of panel *c* in Figure 5.2 is achieved. In this elementary scheme all adjacent atoms are bonded by harmonic interactions. These ones are the sole actions of the model (see Fig. 5.4, panel *a*).

In this case,  $\nu$  reduces to an ellipse, a circle when, starting from a square configuration, the four atoms are driven homogeneously, maintaining the square symmetry. Figure 5.4 displays the evolution in space of the atoms in the square lattice and the corresponding coarse representation in terms of  $\nu$ . The compression phase is not described appropriately because the actual value of  $\nu$  represents an ellipse never smaller than the one corresponding to the initial condition. Also, at the end of a cycle in the simulation, the system is in a state equivalent to the initial one, while  $\nu$  does not reach the initial value.

Previous results on the square lattice refer to a scheme where the reference positions of the atoms, *i.e.* the  $r_i$  in (5.6), are kept constant to the initial values, and the interacting forces  $f_{ij}$  change. However, a similar outcome is detected also in the case the former quantities are let free to evolve, but the two eigenvalues

have different evolutions in magnitude. A similar non-symmetric behavior of the eigenvalues of  $\nu$  occurs also if the number of internal constraints is changed.

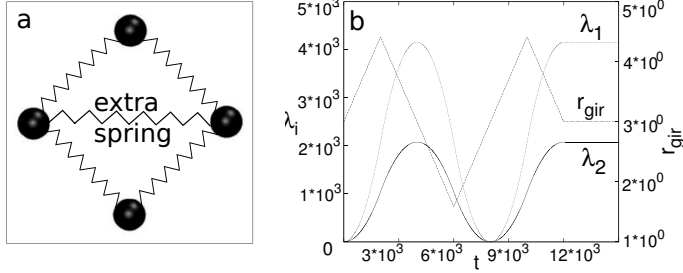


Figure 5.5: Panel *a*: Picture of the four-atoms set where an internal constraint is inserted. Panel *b*: Evolution of the semi axis of the ellipse for a homogeneous deformation totally equivalent to the one depicted in the first row of images of Fig. 5.4. The presence of the extra spring alters the dynamics of  $\nu$ , which semi axes  $\lambda_1$  and  $\lambda_2$  (eigenvalues of the symmetric matrix  $U$  stemming from polar decomposition of  $\nu$ ) follow different trends in spite of the symmetry of the motion.

Indeed, if an extra inner spring is added, as for example the one connecting atoms 2 and 4 in Figure 5.5, panel *a*, but the constant velocity stretching protocol is not varied (the deformation process depicted in the upper row of panels in Fig. 5.4 still holds), we obviously get a different evolution of  $\nu$  which changes from a circle to an ellipse (see Fig. 5.5, panel *b*): the two eigenvalues of  $\nu$  follow, in fact, different dynamics.

So, at least for systems composed by a small number of particles, the expression of  $z_s$  that we have obtained does not allow the description in terms of  $\nu$  to capture appropriately the behavior in compression. However, if we perform the deformation process by using fictitious points, properly connected by springs to the molecule (see the first panel in Fig. 5.6), and we take into account the external forces generated this way, the results appear, at least qualitatively, acceptable. Also, they are qualitatively independent of the number of internal constraints, as summarized in Figure 5.6. Finally, we investigate how the picture changes if we alter the interaction potentials. We enlarge first the number of atoms in the two-dimensional lattice, increasing them up to 8, and consider Lennard-Jones interactions for two different values of the cut-off radius (see Fig. 5.7). For a cut-off  $R_c = 3.0\text{\AA}$  only adjacent residues generate an interaction. At  $R_c = 6.5\text{\AA}$  all atoms interact one another. In Figure 5.7 the system is also depicted in the maximum elongation and compression states, which correspond to labels (1) and (2) in panel *a* of Fig 5.8. Only when the external actions (the ones exerted by the

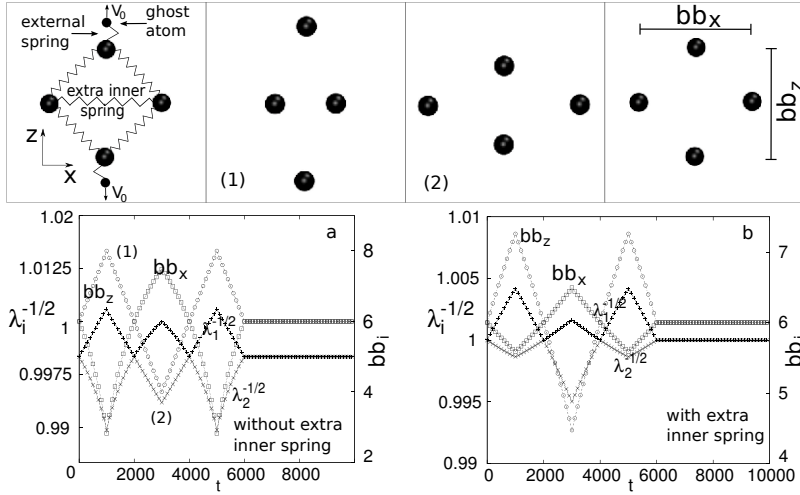


Figure 5.6: First row: Four-atoms system and fictitious beads driving deformation. The ghost atoms follow constant velocity motions and drift the system deformation thanks to harmonic external potentials. Maximum elongation states are also reported and labeled by (1) and (2) also in panel *a*. The overall shape of the system is here summarized by the magnitude of the edges of the limiting rectangle,  $bb_x$  and  $bb_z$ . The semi axes of  $\nu$  should be in agreement with the latter ones. Panel *a*: Comparison between the trends of the limiting box and the ellipse semi axes,  $\lambda_i^{-1/2}$  ( $\lambda_i$  are the eigenvalues of the symmetric part of  $\nu$  by polar decomposition). Here external forces due to the ghost atoms and relative external springs are considered in the computation of  $z_s$ . Instead, the inner extra spring is absent (neither in the system itself nor, accordingly, in the computation of the self-action). The qualitative agreement is considerable. Panel *b*: The qualitative agreement does not change if the inner extra spring is added to the system and to the computation of the self-action, differently from what have observed when external forces are not taken into account in the expression of the self-action (Fig.5.5*b*).

fictitious material points) are included in the computation of  $z_s$ , the variation in time of  $\nu$  can be deemed in agreement with the real shape evolution of the lattice (the square-root of the reciprocals of the eigenvalues of the symmetric matrix  $U$  are again considered).

#### 4.2. Three-dimensional atomic clusters

Previous examples have been developed just to clarify the stage and to pave the way to more intricate three-dimensional simulations, which are, at the end, the realistic setting we have to face. Here we report the results obtained from a simple cubic lattice – two extra atoms are placed in the centers of upper and lower faces to simplify the implementation of constant velocity stretching simulations (the scheme is in Fig. 5.11*a*) – and the protein Ubiquitin (see Fig. 5.1).

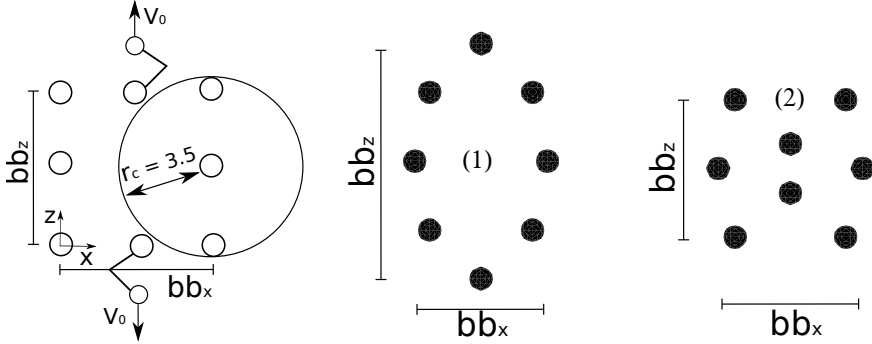


Figure 5.7: 2D eight-atoms system. The smaller cut-off radius used in the simulation is depicted to show that only adjacent points are bonded in this case, similarly to covalent bonding. The external fictitious apparatus and the limiting rectangular box are also easily detectable. Maximum stretching and compression are labeled with marks (1) and (2). The latter ones can be found also in panel *a* of Fig. 5.8 in correspondence with the evolution of  $\nu$  in these states.

**Cubic lattice.** Simulation data arising from the analysis of the cubic lattice sketched in Figure 5.11*a*, confirm what has been shown for the two-dimensional systems.

- The overall landscape depends, of course, on internal interactions among mass points. Passing from short to long-range Lennard-Jones potentials, to covalent harmonic bonding between adjacent atoms, the time evolution of  $\nu$  changes (Fig. 5.9). The results are (in general, indeed) only partially in compliance with the real shape evolution of the lattice (see for example panel *c* in Fig. 5.9).
- The inclusion of the external forces exerted by fictitious atoms, the ones we use to impose deformation in the numerical simulations, is a crucial choice that commonly improves the adherence of the computational results to what we have in mind to describe (see Fig. 5.10). In this setting the square-root of the reciprocals of the eigenvalues of  $U$  appears to be the best way to represent the semi axes of  $\nu$ .
- Thermal fluctuations do not represent a hindrance, providing that lattice stability is not compromised (i.e. the condensed nature of the matter is not altered). The remark does not depend on the possible inclusion of external forces (see panel *b* in Fig. 5.10). However, as evidenced by the results in the same panel (i.e.  $t > 16000t_u$ , with  $t_u$  indicating the time unit used in the computations), since thermal fluctuations are a source of non homogeneous deformation and random forces, in the long run they cause a sort of monotonic trend of the eigenvalues  $\lambda_i$  of  $U$  which is not in compliance with a lattice just vibrating.



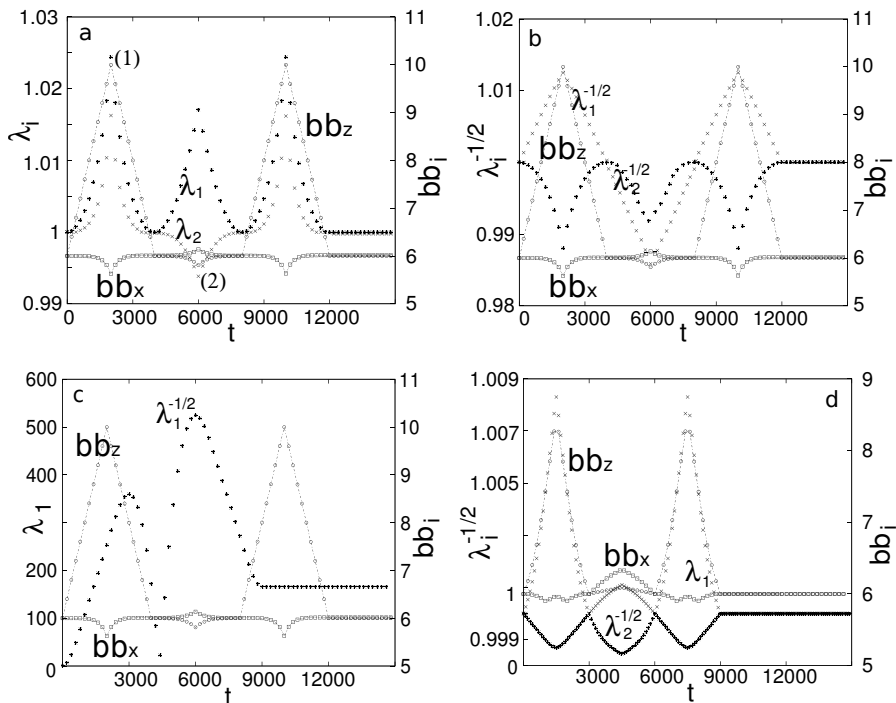


Figure 5.8: 2D eight-atoms system. Panel *a*: Comparison between the evolution of the bounding box edges and  $\nu$  semi-axes. Here the picture refers to a set where short range ( $R_c = 3.5\text{\AA}$ ) Lennard-Jones potentials connect adjacent atoms. The deformation is enforced by external fictitious beads and springs. Such forces are not accounted for in the computation of  $z_s$ . The evolution of semi-axes does not summarize alone the shape of the set of points (there is qualitative agreement only in the compression phase (2)). Panel *b*: Once forces due to external springs are considered, the overall picture improves, providing that  $\lambda_1^{-1/2}$  and  $\lambda_2^{-1/2}$  are taken as the semi-axes of  $\nu$ . Panels *c* and *d*: Similar statements hold true if  $R_c$  is greater ( $6.5\text{\AA}$ ). If external forces are not considered, the evolution of  $\nu$  degenerates into a 1D dynamics (only one of the eigenvalues actually evolves), panel *c*. Again, if we include external forces we recover accuracy and “robustness” in the scheme (panel *d*).

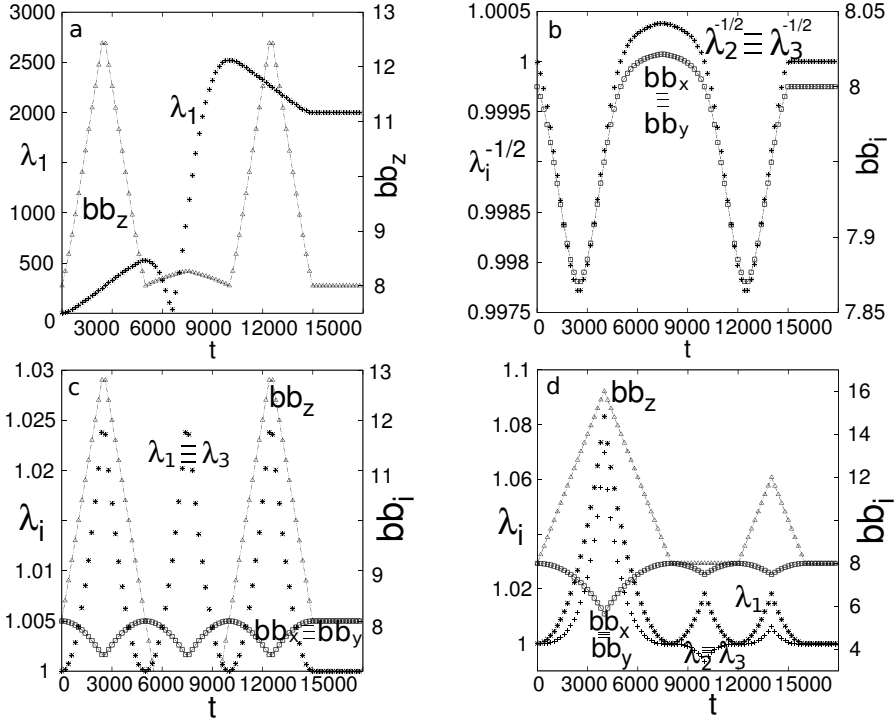


Figure 5.9: Panel *a*: Comparison between the first eigenvalue of  $U$  in (5.2) and the evolution of the edge of the cubic set of atoms oriented along the  $z$ -axis. Stretching and compression are performed along the latter axis by means of ghost atoms and springs, see Fig. 5.11 for snapshots of the system during an equivalent deformation path. There is not a close correlation between the two quantities displayed and a high peak in the evolution of the eigenvalue is generated close to the compression phase, presumably as a consequence of the stiff Lennard-Jones potentials employed in this simulation ( $R_c = 12\text{\AA}$ , all atoms interact with each other). Panel *b*: Trends of second and third eigenvalues of  $\nu$  emerging in the simulation just described. If  $\lambda_1^{-1/2}$  and  $\lambda_2^{-1/2}$  are considered, the qualitative agreement with the edges evolutions is considerable. Unfortunately this approach does not improve the picture of the first eigenvalue. Panel *c*: The same stretching protocol is applied to the system when short range Lennard-Jones interactions are considered ( $R_c = 5.75\text{\AA}$ ). Similarly to the 2D case, short range Lennard-Jones potentials generate an evolution of  $\nu$  which is degenerate: in this case one of the three eigenvalues does not evolve and the picture associated with  $\nu$  is two-dimensional. Also, the trends of the two eigenvalues are equivalent and do not allow to catch which sides of the real shape are shortening and which ones are stretching. Panel *d*: The interaction potentials are in this case harmonic bonds between adjacent atoms of the cubic lattice. The picture of  $\nu$  is correct in the compression phase, that is in the central region of the plot  $8000t_u < t < 13000t_u$ . During the stretching, instead, the three eigenvalues all follow a similar increasing/decreasing trend and none of them allow us to distinguish between which sides of the real shape are shortening and which ones are stretching. Taking the square-root of the reciprocals of all the eigenvalues cannot fix this inadequacy as all trends would be reversed.

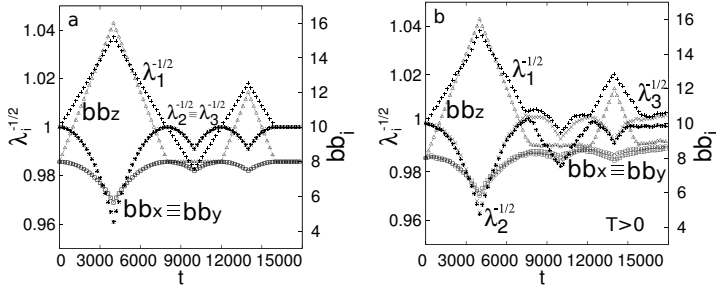


Figure 5.10: Panel *a*: Once external forces due to fictitious atoms and springs that are responsible for system deformation are included in the computation of the self-action, the overall picture results qualitatively in agreement with the dynamics of the three-dimensional cubic set of atoms under scrutiny. The agreement appears when harmonic bonds or short Lennard-Jones interactions are considered, but it is not always the case for high values of  $R_c$ . Panel *b*: Temperature effects. Low-enough values of thermal oscillations (they do not alter the condensed state of the 3D matter) do not affect the time evolution of  $\nu$  during the stretching phases, but generate trends that are not in agreement with the shape evolution of the systems when the external deformation mechanism is no more applied. A slight increasing trend in the values of  $\lambda_1$  and  $\lambda_3$  for  $t > 15000t_u$  can be evidenced, despite no overall deformation occurs.

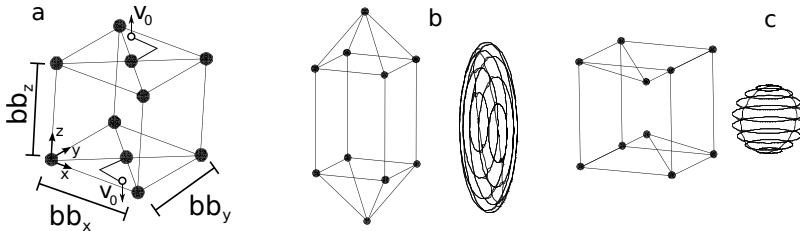


Figure 5.11: Panel *a*: General setting of the cubic lattice. Key quantities such as edge lengths and fictitious atoms, used to perform stretching by assignment of constant velocity motions along the  $z$ -axis, are explicitly depicted. Panels *b* and *c*: Snapshots comparing the evolution of the cubic lattice and the one of  $\nu$  for a system that undergoes extension and compression phases. These images refer to a model in which harmonic potentials bond adjacent atoms and external forces have been used to compute the vectors  $f_{ij}$  in the expression of the self-action. The value of the material element volume  $\epsilon$  has been tuned to achieve comparable trends also from the quantitative point of view.

**Ubiquitin: mechanical stretching.** Ubiquitin protein is composed by 76 residues. The consequent increment of degrees of freedom does not give rise to simulation results differing remarkably in the qualitative features from the ones obtained from the simple lattices considered so far.

In what follows, when we report results in the figures collected in this section, the values of the semi-axes of  $\nu$ , evolving in time, are compared with the smallest rectangular parallelepiped including the protein. Its edges, namely  $bb_1, bb_2, bb_3$ , are evaluated in the central principal frame of the molecular structure (a discussion on the rôle of the coordinate system is also in sec. 4.3). Once again, the approach does not appear to be robust enough to achieve a proper description of the molecular shape unless we introduce (numerically) the external forces exploited in the deformation process, as it happens in the toy schemes discussed previously. However, we have no general proof that the rôle of the external forces in the simulation be always the one we have recognized here.

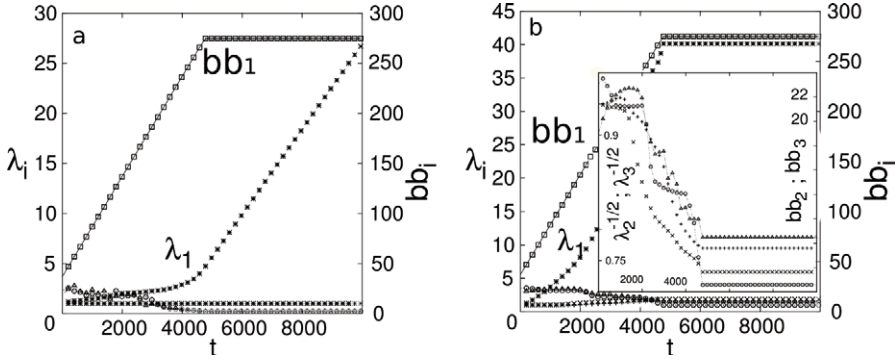


Figure 5.12: Panel *a*: Ubiquitin stretching simulation where the protein is kept elongated at the end of detachment ( $t > 5000t_u$ ). It is evident that the time evolution of  $\nu$  is dominated by the trend of only one eigenvalue. The evolution of  $\nu$  is not in agreement with the one of the molecule shape as  $\lambda_1$  shows a monotonic trend when the molecule is simply oscillating in its linear conformation ( $T = 0$ ). Panel *b*: Once secondary structure terms in the overall potential (3.6) are neglected in the computation of the self action and the temperature is maintained equal to zero, a horizontal plateau is detected in the evolution of the main semi-axis of  $\nu$ . The evolution of  $\lambda_2^{-1/2}$  and  $\lambda_3^{-1/2}$  well describe the trends of the  $x$  and  $y$  axes of the limiting parallelepiped and viewed in the principal central system of inertia of the protein (inset).

Figure 5.12 reports a comparison between two mechanical stretching simulations, where, at the end of detachment, the protein is kept elongated. To obtain an appropriate qualitative description of the molecular shape, it is necessary to neglect the contributions of the secondary structure in the computation of the vectors  $f_{ij}$  that determine  $z_s$ , and to set the temperature equal to zero (see panel

$b$  in Fig. 5.12). The need appears, obviously, because the elongated configuration is not an equilibrium one. Internal forces do not vanish – the protein oscillates driven by the ground state of the potential (3.6) – hence the self-action does not vanish and determines the evolution of  $\nu$  (see panel  $a$  in Fig. 5.12). Only when the contribute of such forces is neglected we obtain what is depicted in panel  $b$  of the same figure. We just notice that the results improve if we consider the square-root of the reciprocals of the second and third eigenvalues, differently from the first one, although there is no evident physical reason to justify such a choice. The picture holds true providing that peptide bonds are not loaded, otherwise the horizontal plateau in the trend of  $\lambda_1$  would be impossible. Thermal fluctuations also compromise the evolution of  $\nu$  in a similar way. Also, such a shortcoming cannot be easily eliminated by including the external forces by which we generate numerically the deformation path, as they would simply add a contribute to two of the vectors  $f_{ij}$ , which does not guarantee that  $z_s$  would vanish. Actually there is only one eigenvalue of  $U$  that evolves appreciably and governs the evolution of  $\nu$ , despite the analysis be three-dimensional.

In analyzing the whole protein, to obtain a clear picture of the evolution of the molecular shape it is not always sufficient to pull just one end of it and include the external force in the computation of  $z_s$ . As it appears in comparing panels  $a$  and  $b$  in Figure 5.13, only when both ends are pulled far apart from each other and the two external forces – the ones due to the springs that drag the molecule ends – are included together in the scheme (panel  $b$ ), the picture is qualitatively in agreement with the real evolution of the protein shape (and it seems ‘robust’ as indicated by the additional examples in sec. 4.3). Figure 5.14 compares the shape of the Ubiquitin with its coarse approximation in terms of  $\nu$ . The approximation process based on the ideas underlying Cauchy-Born rule can be optimized by varying the amplitude of the normalization volume  $\epsilon$ . We can choose it, in fact, in order to achieve comparable amplitudes between the limiting parallelepiped and  $\nu$ , with the possible need of a smoothing algorithm in post-processing.

The effect of temperature is usually negligible, during stretching and in the phase immediately consequent, but it influences the dynamics of  $\nu$  in the long run, as we have already affirmed for the toy models analyzed previously. In the scenario described in this section, changes in the cut-off radius, frame of reference or stretching velocity do not induce prominent qualitative alterations.

### 4.3. Additional remarks

As we have repeatedly affirmed, the continuum approach proposed here, with the specification of structure of  $z_s$  in terms of the actions in the discrete model, seems to capture the dynamics of a single macromolecule in presence of constraints imposed by the environment. So, it can be used in representing the mechanical stretching of a protein in free space. For results relevant to translocation dynamics in terms of the continuum model proposed here, see section 4.5.

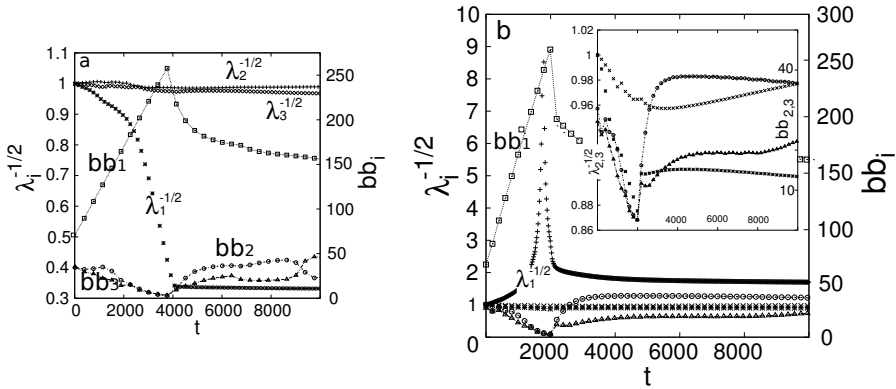


Figure 5.13: Panel *a*: Data refer to a mechanical stretching simulation in which one end of the molecule is pulled at constant velocity by a fictitious bead and an external spring and the other end is held fixed in space. Despite the external force of the spring is included in the computation of the self-action, its presence is not enough to achieve an adequate representation of the molecule shape. Panel *b*: Only when mechanical stretching is performed pulling both ends with equal and opposite velocities, including both the two external spring forces in the self-action, the behavior in time of  $\nu$  is in agreement with the real shape evolution of the protein. It is, as usual, necessary to take the square-root of the reciprocals of the eigenvalues as  $\nu$  semi axes.

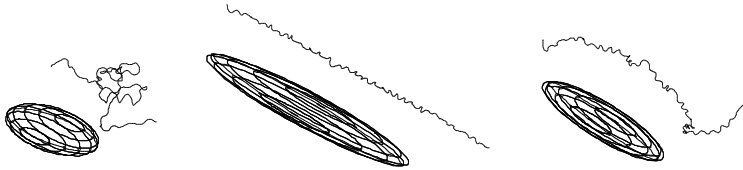


Figure 5.14: Quantitative comparison between discrete Ubiquitin dynamics and the evolution of  $\nu$  when both ends are pulled by external atoms. The material element volume has been optimized to allow the quantitative good agreement. Temperature is set equal to 0. Panel *a*: Beginning of mechanical pulling. Panel *b*: Maximum stretching state. Panel *c*: Conformation of the protein after maximum detachment has been achieved and the protein is left free to refold onto its native structure.

The scheme should be further enriched in case we would consider a deforming environment and/or the presence of a dense population of interacting macromolecules. In the latter case the balance of actions (5.5) should include a microstress, due to the interactions of each molecule with the neighboring ones. Moreover,  $\nu$  should not necessarily be associated with a single molecule, rather it could be a representative ‘in average’ of a number of molecules occupying a portion of space considered as material element and assigned to a point in a multi-

scale and multi-field continuum representation of the macromolecular population (this way we should go along the guidelines of the mechanics of complex materials [3]).

In any case, the inadequacy in presence of molecular free motion remains, as it will be shown in the conclusive part of the present section.

It is possible to overcome it by adopting a second-rank tensor  $\nu$  as a descriptor of the single molecule with a different meaning from the one assigned here, and selecting an expression for  $z_s$  different from (5.6) but *inspired* by it. Possible approaches and related results shall be discussed in section 5.

**Effect of time-dependent placement vectors for Ubiquitin.** Figure 5.15 displays the trends of the semi axes of the ellipsoid associated with  $U$  when placement vectors in the explicit expression of the self action  $z_s$  depend on time, meaning they are computed at every time step on the actual configuration of the protein,  $r_i = r_i(t)$ . It is clear that the monotonic behavior, theoretically predicted from considerations on 1D systems, still holds for greater sets of atoms.

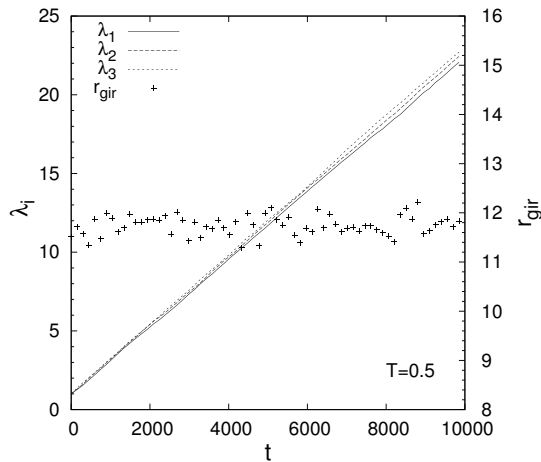


Figure 5.15: Trends of the eigenvalues of  $U$  in a thermalization simulation for the Ubiquitin protein when placement vectors  $r_i$  in  $z_s$  vary in time. The behavior displayed by the analytical results for the biatomic system are confirmed also for the Ubiquitin protein. Data refer to an equilibration temperature  $T = 0.5$ . The molecule maintains its folded state: the radius of gyration undergoes just small fluctuations around a constant value. The semi-axes of the morphological descriptor follow a monotonic evolution not in accordance with the protein shape dynamics.

**Frame of Reference.** A few remarks on the way we have used frames of reference in computations are collected here.

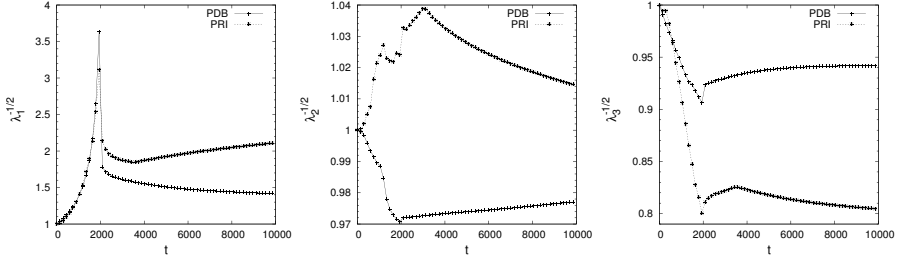


Figure 5.16: Comparison between  $U$  semi-axes for stretching runs relevant to the PDB frame and to the central principal system (PRI) for the Ubiquitin protein. Evolutions can be considered equivalent, despite an opposite trend for the intermediate eigenvalue that, however, results nugatory for the final shape of the morphological descriptor. Data refer to a simulation where both terminuses are dragged with opposite velocities,  $T = 0$ , and external fictitious forces are included in the computation of  $z_s$ .

When we consider just the reference values of the placement vectors  $r_i$ , under rotations  $R \in SO(3)$  of the ambient space (where we evaluate the vectors  $f_{ij}$ ) alone,  $z_s$  becomes  $Rz_s$  because the generic addendum  $f_{ij} \otimes r_i$  defining it changes into  $Rf_{ij} \otimes r_i$ , for  $r_i$  is fixed once and for all in the reference space that is different from the ambient one and isomorphic to it, the isomorphism being simply the identification. In principle, we could also consider rotations  $R' \in SO(3)$  in the reference space so that  $z_s$  would change into  $Rz_s R'^T$  with the possibility of having  $Rz_s R^T$  when  $R' = R$ . Tensor  $\nu$  follows analogous rules, being the integral in time of negative  $z_s$ , according to the evolution equation  $\dot{\nu} = -z_s$ .

In contrast, when we take  $r_i$  in the current configuration, as in the previous example, under rotations  $R$  of the ambient space  $z_s$  is mapped into  $Rz_s R^T$ , and  $\nu$ , evaluated this time in the current shape of the molecule, undergoes an analogous change.

Consequently, when we compute  $\nu$  in one frame and want to know it in another one rotating in time, we have to know the whole time-history of the rotation up to the instant considered.

Here we have considered the frame of reference associated to the protein data bank file (PDB), which is the one where protein coordinates are provided and can be selected once for all at will in space. However, if we develop computations with respect to the principal inertial frame (PRI) of the molecule, the results do not alter the picture that we have shown in previous sections. Figure 5.16 shows that the overall shape of the molecular descriptor is equivalent in both cases, in the present example, if quantities are computed with respect to the protein data bank reference or to the central principal frame, despite an opposite trend in  $\lambda_2^{-1/2}$ . Broadly speaking, on average, differences in between the two cases can be more remarkable than what reported here, nevertheless a final argument to prefer the central principal system to the PDB one is not apparent from our analyses.



## Coarse-grained molecular dynamics and continuum models

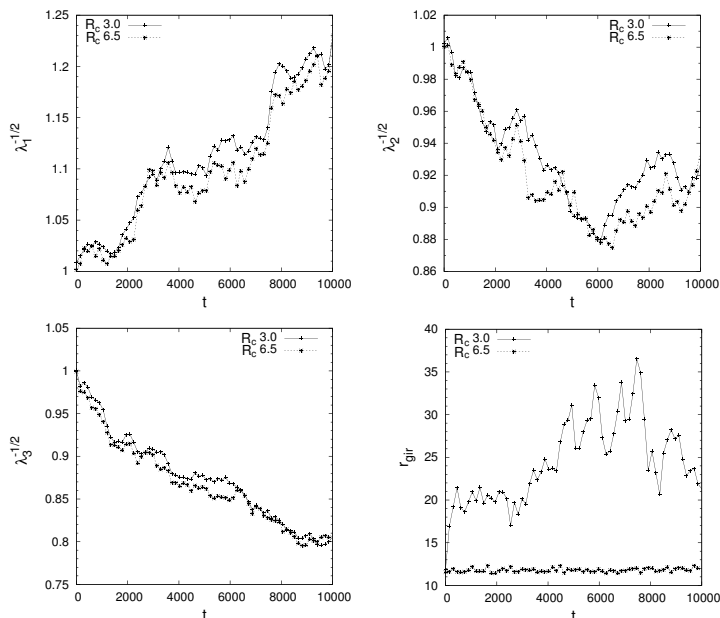


Figure 5.17: Trends of the square root of the reciprocal of the eigenvalues of  $U$  in a thermalization simulation of the Ubiquitin protein.  $R_c = 3.0\text{\AA}$  and  $R_c = 6.5\text{\AA}$ . The first three pictures show that the evolutions are practically the same. However, the protein with  $R_c = 3.0\text{\AA}$  is unfolding, according to the comparison depicted in the last panel, where the trends of the radii of gyration of the two molecules is showed. The approach followed here does not allow to correctly detect thermal unfolding, a situation dominated by inhomogeneous deformation and random fluctuations that do not contribute to the overall shape evolution.

**Thermalization simulations.** For each cut-off radius considered in previous analyses we have performed thermalization simulations at  $T = 0.50\theta$  (in the already mentioned code units) to measure the evolution of the molecular descriptor associated with the protein shape. The protein is free to fluctuate in the ambient space, following the Langevin dynamics already mentioned, with no driving forces nor additional constraints. Each simulation starts from the crystallized structure of the protein according to the PDB file. Unfolding is then allowed, providing that the value of the cut-off radius be small enough. Protein configurations are also regularly sampled to obtain statistical independent initial states for mechanical stretching simulations. As it was expected, changes in the molecular shape, due to thermal equilibration, are not properly detected by the approach outlined in this work, since the deformation process is inhomogeneous and not remarkable (the radius of gyration of a thermally unfolded protein is only about two times greater than the one own of a compact structure). In addition, single-residue fluctuations play a prominent rôle at the atomic level, with-

out a corresponding macroscopic deformation. At the given equilibration temperature, structures with cut-off radius equal to  $3.0\text{\AA}$  unfold, while if  $R_c = 6.5\text{\AA}$  they do not. Figure 5.17 depicts these remarks for the Ubiquitin protein. The trends of the semi-axes of  $U$  are similar and mostly indistinguishable, despite a protein undergoes thermal unfolding (as it results from the trends in the radius of gyration). A similar behavior is found in all the equilibration simulations that we have performed.

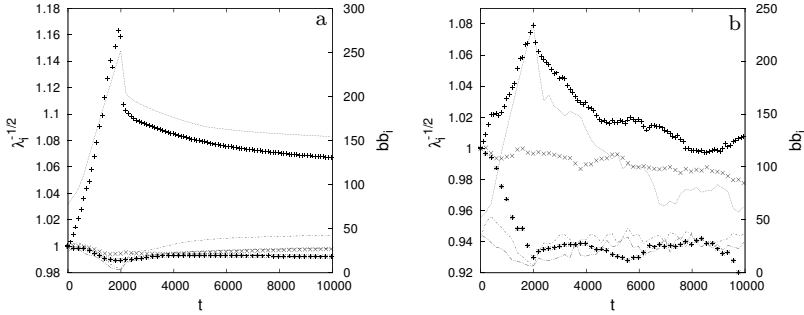


Figure 5.18: Panel *a*: Trends of the reciprocals of the square root of the eigenvalues of  $U$  in a stretching simulation of Ubiquitin compared with the dynamics of the smallest rectangular parallelepiped containing the molecule. The eigenvalues are here depicted as sequences of points, the values of the edge of the parallelepiped as continuous lines.  $R_c = 3.0\text{\AA}$ ,  $T = 0$ ,  $v_{str} = 0.05$ . Data refer to a simulation performed with respect to the principal central frame. Panel *b*: Analogous data than in panel *a* but  $T = 0.2$ . Molecular descriptor data are computed with respect to the PDB system of coordinates.

**Additional results for stretching simulations.** We present here some additional data which can indicate the statistical robustness of the results provided above, concerning the free-space stretching of the protein Ubiquitin when both ends are pulled at opposite constant velocities and external forces pertinent to the fictitious atoms are included in the computation of  $z_s$ , see Figures 5.18-5.20. The material element volume  $\epsilon$  has not been optimized for quantitative comparisons for the data below, as here we are just interested in furnishing a qualitative picture. Specifically, we have performed ten independent simulations for each cut-off radius ( $R_c$ ), temperature ( $T$ ), and stretching velocity ( $v_{str}$ ) implemented ( $R_c = 3.0\text{\AA}$  or  $R_c = 6.5\text{\AA}$ ,  $T = 0$  or  $T = 0.2$ ,  $v_{str} = 0.05$  or  $v_{str} = 0.1$ ). For all the cases, results can be deemed qualitatively equivalent to the few additional examples here provided.

## Coarse-grained molecular dynamics and continuum models

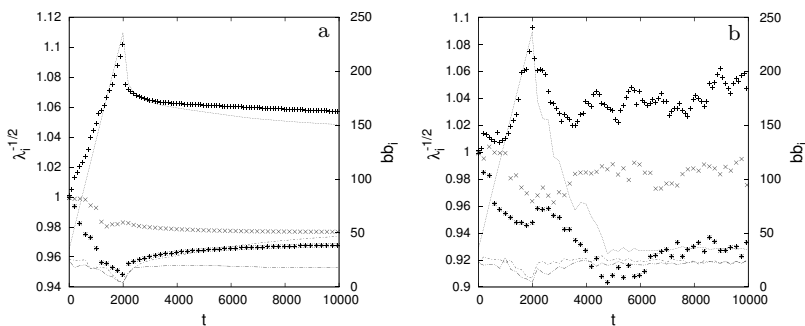


Figure 5.19: Panel *a*: Trends of the reciprocals of the square root of the eigenvalues of  $U$  in a stretching simulation of Ubiquitin compared with the dynamics of the smallest rectangular parallelepiped containing the molecule. The eigenvalues are here depicted as sequences of points, the values of the edge of the parallelepiped as continuous lines.  $R_c = 6.5\text{\AA}$ ,  $T = 0$ ,  $v_{str} = 0.05$ . Data refer to a simulation performed with respect to the principal central frame. Panel *b*: Analogous data than in panel *a* but  $T = 0.2$ . Molecular descriptor data are computed with respect to the PDB system of coordinates.

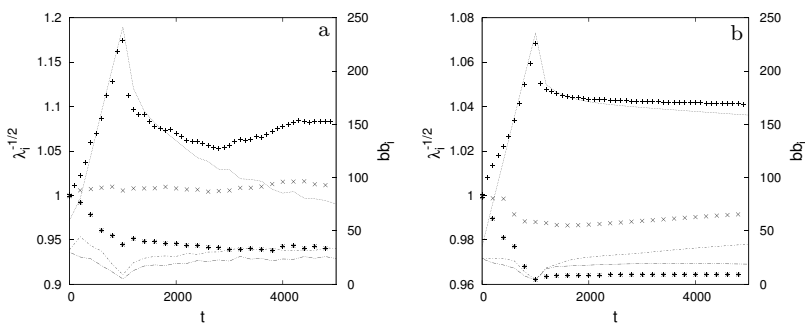


Figure 5.20: Panel *a*: Trends of the average over 10 runs of the reciprocals of the square root of the eigenvalues of  $U$  in stretching simulations of Ubiquitin (data relevant to the central principal frame). These quantities are compared with the average dynamics of the bounding box. The eigenvalues are here depicted as sequences of points, the values of the edges of the parallelepiped as continuous lines.  $R_c = 3.0\text{\AA}$ ,  $T = 0.2$ ,  $v_{str} = 0.1$ . Panel *b*: Analogous average data than in panel *a* but now  $R_c = 6.5\text{\AA}$ ,  $T = 0.0$ ,  $v_{str} = 0.1$ . Molecular descriptor data are here computed with respect to the PDB system of coordinates.

### 4.4. Protein descriptor in MBP stretching simulations

As already mentioned for the Ubiquitin protein, placement vectors  $r_i$  refer to reference (equilibrated) configuration and do not depend on time. Here results relevant to the first stretching protocol are presented. In fact for the MBP, in contrast to what shown previously, we find it sufficient to pull just one end of the molecule (keeping the other terminus blocked in space) to gain an appropriate

picture of protein shape evolution. As usual, the external force generated by the ghost atom that drives the deformation pathway is included in the computation of the self-action and the square root of the reciprocals of the eigenvalues of  $U$  are again depicted as the semi axes of an ellipsoid.

In what reported in Figure 5.21 (N-terminus pulling,  $R_c = 7.5$ ,  $T = 0.75$ ,  $v_{str} = 0.1/t_u$ ), the material element volume  $|e|$  has been optimized for quantitative comparisons. In this case raw data are worked out to gain the trends and images from panel *a*2 to panel *f*. First, original evolutions are smoothed by supposing a process with memory. Namely a linear-weighted average over 1000 previous values is performed for each data. This way trends can be softened and, especially,  $\lambda_1^{-1/2}$  (the one that practically governs the description) is closer to  $bb_1$  evolution, also in the equilibration part. Panels *a*2, *b*2 and *c*2 report the outcome of the application of this algorithm to the  $\lambda_i^{-1/2}$  trends. Finally, to achieve the ellipsoid snapshots only (panels *d*, *e* and *f*) a magnification coefficient equal to 4 is used to recover a high-enough value of the first peak, as it can be noted that it is considerably reduced by the smoothing. No magnifications or reductions are used for  $\lambda_2^{-1/2}$  and  $\lambda_3^{-1/2}$ , as the average trend of those is deemed appropriate. To get the pictures, the center of the ellipsoid is conveniently attached to the center of mass of the molecule and its orientation is governed by the eigenvectors of  $U$ . Thanks to these expedients, the description appears appropriate, despite the evolution of the medium and smaller axes may be deeper.

We add that in the present approach the equivalent radius of the sphere having the same volume of the ellipsoid follows a trend dominated by the first eigenvalue, that is a trend qualitatively equivalent to the radius of gyration one (the volume increases in a stretching simulation).

#### 4.5. MBP morphological descriptor translocation results

In translocation simulations we find that the presence of the pore, which constitutes a constraint that limits accessibility in phase space, enhances the quality of the protein shape description. To the sake of clarity, in the present approach, both the pulling force and the actions, due to the channel potential, are included in the computation of the  $z_s$  as external forces. The confinement acts similarly to a decrement of the temperature, limiting residue fluctuations inside the pore. Obviously, to achieve a proper outcome, either the volume of the material element  $|e|$  needs to be small or the pore long enough to provoke a considerable stretching, detectable by the eigenvalues of  $U$ . The issue is well summarized in Figure 5.22, where results related to two different pore lengths are depicted, holding the material element volume constant at a standard value (equivalent sphere radius:  $30\text{\AA}$ ). Thus, once again, main limitations are due to the necessity to fine-tune such a volume on the basis of the simulation performed, or rather, according to a sort of ‘maximum value of deformation’.

## Coarse-grained molecular dynamics and continuum models

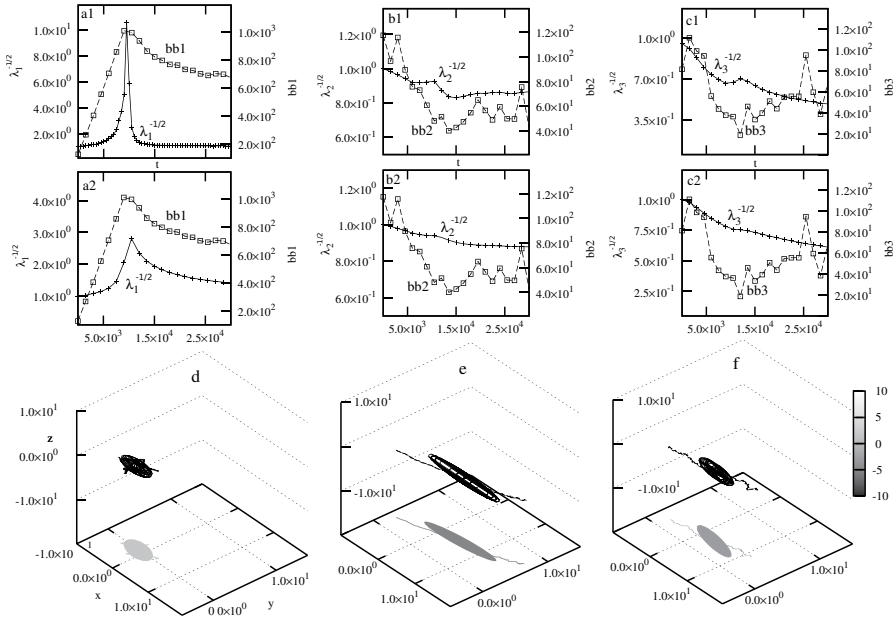


Figure 5.21: Raw and elaborated data are shown for a stretching simulation of Maltose Binding protein. As evident from the first row, to gain a better agreement between peak magnitudes,  $|\epsilon|$  should be further reduced, but this would generate instabilities in the routines used for polar decomposition. When depicted, protein dimensions are normalized by the radius of gyration own of the PDB crystallized structure. Panel *a1*: Trend of the original  $\lambda_1^{-1/2}$  as resulting from fine tuning of  $|\epsilon|$ . The peak is high, reaching a value eleventh times greater than the initial one. Its evolution is also tight and sharply decreases during thermalization. Panel *b1*: Despite the reduction in the value of  $|\epsilon|$  the drop at maximum detachment does not overcome 10% of the initial value and is equal to 20% at the end of the simulation: the dull  $\lambda_2^{-1/2}$  dynamics is a limitation of the model. Panel *c1*: The minimum value at  $10000t_u$  corresponds to a drop of 30% of the initial value and has to be compared to a 70% reduction in  $bb_3$ , so the agreement is not very close in magnitude. Panel *a2*:  $\lambda_1^{-1/2}$  trend resulting from a weighted running average. Specifically, the weights are linear (from 0 to 1) and spread over the current value (weight = 1) back to the one-thousandth previous one (weight = 0). The approach allows improving the agreement between  $\lambda_1^{-1/2}$  and  $bb_1$ , especially in the thermalization part. Panel *b2* and *c2*: Same approach just described applied to  $\lambda_2^{-1/2}$  and  $\lambda_3^{-1/2}$ . A similar smoother behavior is detected and amplitudes are reduced. Panel *d*: Snapshot of the ellipsoid and molecule at the beginning of the pulling. The protein is compact and the ellipsoid is close to a sphere. Here and in panels *e* and *f* a magnification coefficient is applied to the ellipsoid semi axis to recover the amplitude of the peak smoothed by the averaging process. Panel *e*: Ellipsoid and protein near the maximum value of the stretching. A remarkable elongation is detected in agreement with the almost linear fashion of the polymer. Panel *f*: Snapshot taken at the end of equilibration. The morphological ellipsoid shrinks still too quickly from its maximum elongation when compared to its shape in panel *d*.

Panel *a*, *b* and *c* underline the difference in the amplitudes of the square root of the reciprocals of the eigenvalues once the channel is elongated from 100Å to 600Å. Panels *d* through *f* compare the MBP shape and the ellipsoid associated with  $U$  once projected onto the plane formed by eigenvectors 1 and 3 (600Å channel). The choice of the plane is not by chance, as again we observe that eigenvalues 1 and 3 govern the dynamics, while the second one experiences just little fluctuations. The outlined approach confirms its appropriateness to summarize shape evolutions for a wide range of simulation parameters in constrained dynamics. We performed several transport simulations across a nanopore by varying the cut-off radius, pore length, temperature, driving force (always considerably great, to perform translocation in a limited time window) and pulling terminus, without remarkable limitations in the description (providing that the transport be homogeneous in time).

However, the model is sensitive to the value of the pulling force: a greater value causes higher values of interatomic forces  $f_{ij}$ , which are detected by  $z_s$ . Hence, two translocation simulations, which are equivalent from a macroscopic deformation point of view (same channel), just performed at different driving forces, are described by different  $\nu$ . The one relevant to the greater value of the pulling force generates the ellipsoid that stretches the most, whereas instead the gross deformation of the proteins is practically the same. A similar behavior is also noticed in stretching runs when detachment velocity is increased.

## 5. Additional continuum formulations

As we have seen, the proposed approach based on Cauchy-Born rule and power equivalence, appears to be robust enough only when external forces are included in the computation of  $z_s$ . Also, the dynamics has to be, in a sense, driven, constrained, by the external environment. However, additional preliminary results (not shown here since they are deemed unreliable as a proper statistics of event has not been collected yet) seem to show that external forces might not be essential for MBP stretching and translocation simulations. In other words the effect of the number of degrees of freedom of the system might require further investigations.

What appears evident are the limitations of the model to capture thermal unfolding. In thermal simulations the homogeneity is lost due to random single-bead oscillations that overwhelm and hide the overall unfolding process. The drawback can be overcome by selecting different descriptors of the protein evolution and related modified expressions of  $z_s$ , which arise from evaluations of the interaction mechanisms among material points in the molecule.

## Coarse-grained molecular dynamics and continuum models

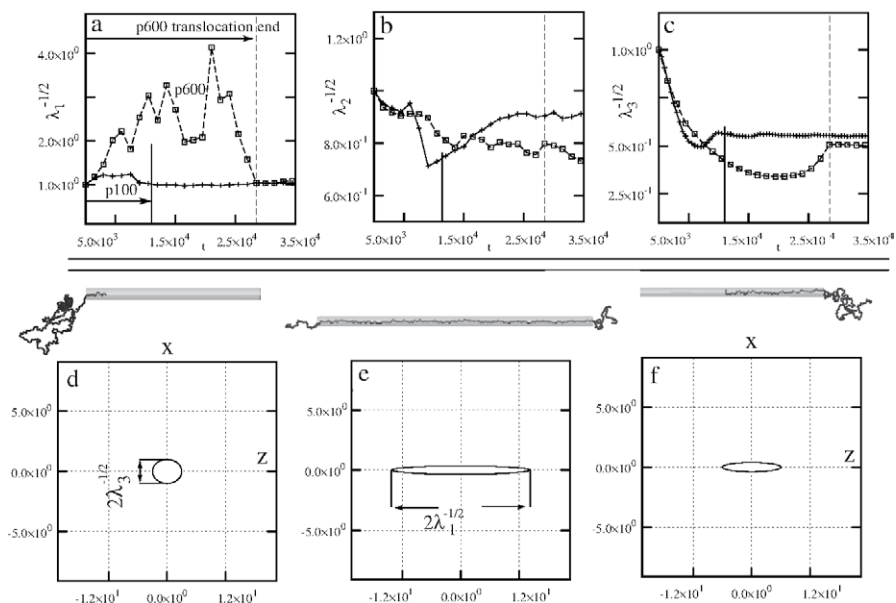


Figure 5.22: Translocation simulations for  $R_c = 3.0$ , C-terminus pulling,  $T = 0.75$ ,  $F = 2.0$ . Panels *a*, *b* and *c*: Trends of the reciprocals of the eigenvalues for two values of the pore length (100 and 600Å). Despite a lower overall deformation if compared with the stretching simulations, the presence of the pore allows a better description of the molecule shape evolution. Especially  $\lambda_3^{-1/2}$  (panel *c*) appears to be particularly sensitive to the confinement, its decreasing trend is considerable, reaching a minimum equal to about one half of the initial value, regardless the pore length.  $\lambda_1^{-1/2}$  is instead influenced by the latter quantity (panel *a*), being negligible in the 100Å pore. Finally, the evolution of the second eigenvalue, although again pretty limited, is remarkable and usually greater than in the steered-dynamics case. Panels *d*, *e* and *f*: Direct comparison between the ellipsoid projected onto the planes associated with eigenvectors 1 and 3 and the molecule seen in the  $xz$  plane of the PDB system of coordinates. The quality of the description is considerable regardless the frame of reference employed. Due to the simulation protocol (alignment of the PBD x-axis with the pore axis and therefore with the pulling force direction), all the frames are very close to each other.

### 5.1. Varying the meaning of $\nu$

We change here the meaning of  $\nu$  and consider it as a *work tensor density* performed ‘in average’ in deforming the whole protein. Also here we assume  $\det(\nu) > 0$ . This way  $\dot{\nu}$  is a power and it is supposed to be equal to the power density in the discrete model. The equation  $\nu = -z_s$  is now a *tensor power balance*. The structure of  $z_s$ , here indicated by  $z_{s2}$ , is that of a tensor power density. We express (heuristically)  $z_{s2}$  as the sum of the tensorial product between the absolute value of the total force  $|f_{ij}|$  exerted between residues  $i$  and  $j$  and the time rate of change of the absolute value of the placement vector of the material

points with respect to the center of mass of the molecule,  $\partial_t|r_i|$ . It is important to point out that here  $|\cdot|$  indicates the absolute value of the components of the considered quantity, that is  $|x|$  is a vector which entries are the absolute values of the entries of vector  $x$ , it is not its norm, which in turn would be indicated by the symbol  $\|x\|$ . In a sense the dynamics here is like projected onto the first octant of the Cartesian space  $\mathbf{R}^3$ , through proper reflections (associated with the absolute value operator) that allow us to distinguish between departure or approach of a residue with respect to the center of mass, regardless their mutual placement in the frame of reference selected to assign point-coordinates. We write

$$z_{s2} = -\frac{1}{|\mathbf{e}|} \sum_{i \neq j \in \mathbf{e}} k_i^{z_{s2}} |f_{ij}| \otimes \partial_t|r_i|. \quad (5.7)$$

In the above expression  $k_i^{z_{s2}}$  is a dimensionless scalar quantity defined as the ratio between the actual distance of the point of application of the force from the center of mass of the molecule,  $|r_i|$ , and the radius of gyration of the crystallized PDB structure of the protein,  $r_{gir}^0$ :

$$k_i^{z_{s2}} = \frac{\|r_i\|}{r_{gir}^0}.$$

The action  $f_{ij}$  involves all the elementary terms present in the Gō-like model, such as actions stemming from peptide bonds, bending deformation and so on, equivalently to the previous approach. As evident from (5.7), not only the square root of the reciprocal of the eigenvalues of the symmetric matrix due to polar decomposition of  $\nu$  are analyzed, but also the diagonal entries of  $\nu$  are investigated and a suited ellipsoid built from them.

**Thermalization simulations results.** For a matter of brevity, we show only thermalization simulation results. In fact only in this case the first approach considered in the present chapter is not appropriate. However, similar outcomes (in terms of the quality of the overall picture) hold also for stretching and translocation simulations. Nevertheless, the description of thermal unfolding is the toughest task. The approach analyzed here is mainly tailored to follow protein deformation when diagonal components of  $\nu$  are considered. However, also symmetric matrix eigenvalues provide an adequate geometrical description of the molecule.

In Figure 5.23 we show the trends of  $\nu$  diagonal entries for a molecule with cut-off radius equal to  $7.5\text{\AA}$ , while undergoes equilibration. As already shown (Fig. 3.5), since the simulation temperature is 0.75, MBP remains compact. Also, that behavior can be easily detected by the lack of increment in the bounding box edge lengths, as they simply oscillate around constant values regardless the frame of reference used (all panels of Fig. 5.23). In panels *a*, *b* and *c* we compare the



trends of the square root of the reciprocals of the eigenvalues of  $U$  with the evolution of the sides of the limiting rectangular parallelepiped in the central principal frame, despite the simulation runs with respect to the PDB system of coordinate. We do this because we believe that the description by means of symmetric matrix eigenvalues should be more general than the one based on the diagonal entries of  $\nu$ . All the panels show a good agreement between the two trends reported in each of them. The eigenvalues do not show the almost monotonic trends as the ones in Figure 5.17. Rather, they randomly fluctuate around the initial values, a behavior in accord with the bounding box one. However, fast spikes are not reproduced and sometimes a sort of opposite behavior is detected. Panels *d*, *e* and *f* report the evolution of  $\nu$  diagonal entries and bounding box edges in the PDB frame of reference ( $bb_x, bb_y, bb_z$ ).  $\nu$  diagonal entries are indeed tied to the system of coordinates where quantities are evaluated. It is evident how much these trends faithfully reproduce the evolution of the molecule shape even in such a random dynamics. The magnitude of the oscillations can be easily changed by altering the value of the material element volume once the approach here analyzed is chosen. Fast spikes are closely matched. Such an agreement holds true for all the simulations we have performed, regardless the value of temperature and cut-off radius. Therefore the associated ellipsoid always describes the protein shape in a pointwise fashion, not only in average. Unfortunately, neither the eigenvectors of  $\nu$  nor of  $U$ , (not even the rotation matrix arising from polar decomposition) provide a straightforward way to understand the orientation of the polymer. Also, in this particular case, since the molecule remains folded,  $\nu$  diagonal components are really able to follow the details of the deformation pathway (bounding box oscillations are only a few Angstroms), which is not macroscopic. To understand what just stated, see Figure 5.24. It clearly shows the ability of the current approach to distinguish between a protein that remains compact in a heat reservoir and one that instead unfolds. Indeed, panels *a* through *c* compare three quantities: square root of the reciprocals of the eigenvalues of  $U$  for a 3.0Å cut-off radius molecule, the relevant evolution of the bounding box sides and the square root of the reciprocals of the eigenvalues of  $U$  for a protein that remains folded. First of all we underline that the agreement between eigenvalues and associated edges of the limiting rectangular parallelepiped is appropriate on an average sense, since bounding box spikes are not so closely followed by ellipsoid semi-axes. However, the magnitude of the deformation is clearly detected: fluctuations in the evolution of  $\lambda_i^{-1/2}$  associated with the folded molecule are hardly visible. Such quantities remain practically flat and very close to the initial value in a macroscopic standard, that is the one determined by the fluctuations belonging to the 3.0Å cut-off radius MBP. It is evident that we can recognize that the latter one is undergoing big deformations (thermal unfolding), while the former one practically remains in a native-like conformation. Also, now it is clear what we meant about the ability of the selected approach to follow the little details of the deformation process, even the smallest changes in the shape are

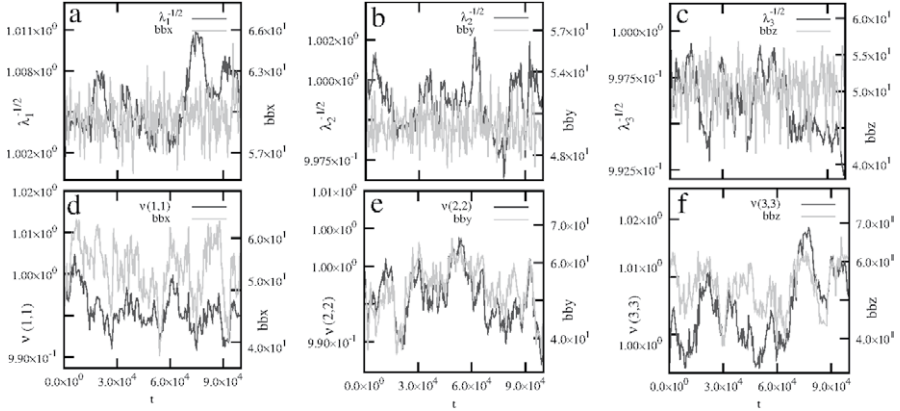


Figure 5.23: Thermalization simulation results for MBP.  $R_c = 7.5$ ,  $T = 0.75$ . The protein remains folded with these parameters. Panels *a*, *b* and *c*: Comparison between the bounding box side lengths and the square root of the reciprocals of the eigenvalues of  $U$ . The former are computed in the principal central frame, the latter are relevant to the polar decomposition of  $\nu$  evaluated in the PDB coordinate system. In panel *a* the highest-valued eigenvalue trend is shown, in panel *b* the middle-valued one and, accordingly, in panel *c* the lowest. All three show a remarkable agreement with the trends of the bounding box edges, especially considering that are relevant to different frames, in compliance with the need that the eigenvalues of the symmetric matrix should have a general meaning. Panels *d*, *e* and *f*: Comparison of  $\nu$  diagonal components and bounding box sides in the PDB frame. The agreement is in all cases at pointwise level, spikes are correctly reproduced, despite the gross deformation of the molecule is negligible as no thermal unfolding takes place.

recorded and are scaled correctly, providing that the material element volume of the compared evolutions be the same.

The same picture is confirmed in panels *d*, *e* and *f* of Figure 5.24 where diagonal components are considered. Also in this case unfolding is clearly detected. Moreover, the agreement between the ellipsoid semi axes and the bounding box sides appears adequate not only on average but pointwise. If the ellipsoids are plotted, the difference in their dimensions is evident and we can detect which molecule is unfolding, as shown in Figure 5.25. Data used for this figure are the ones also reported in Figure 5.24, when the difference between the first eigenvalues is close to the maximum one.

## 5.2. $z_s$ model 3 and 3\*

**Explicit formulation.** We can also choose different expressions of  $z_s$  as explicit constitutive structure. Two of them are discussed here. The first one, indicated

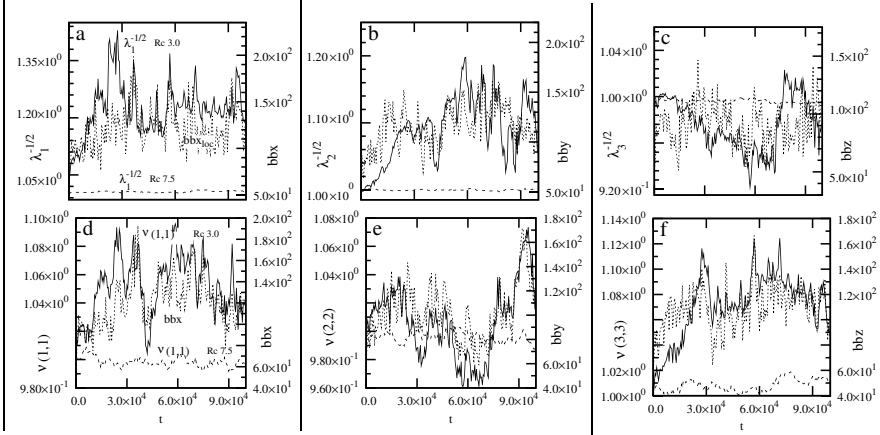


Figure 5.24: Thermalization simulation results for MBP.  $R_c = 3.0\text{\AA}$ , and  $R_c = 7.5\text{\AA}$ ,  $T = 0.75$ . The molecule with the lower  $R_c$  unfolds, the other one remains compact. The material element volume is the same for the two cases, equal to a sphere of radius  $20\text{\AA}$ . Panels *a*, *b* and *c*: Comparison between the bounding box side lengths (solid dark line) and the reciprocals of the eigenvalues of the  $3.0\text{\AA}$   $R_c$  molecule (dotted line). Also the reciprocals of the eigenvalues of the  $7.5\text{\AA}$   $R_c$  protein are shown for a quantitative comparison (dashed line). The agreement between the  $3.0\text{\AA}$  case and relevant bounding box sides is appropriate in all panels. Also, the current approach is able to distinguish between a molecule that undergoes thermal unfolding and one that does not, accordingly to the difference in  $\lambda_i^{-1/2}$  magnitude. Panels *d*, *e* and *f*: Comparison of  $\nu$  diagonal elements and limiting parallelepiped edges in the PDB frame of reference for the  $R_c = 3.0\text{\AA}$  simulation. The agreement is appropriate in all cases and even fast spikes are closely followed.

by  $z_{s3}$ , reads

$$z_{s3} = -\frac{1}{|\mathbf{c}|} \sum_{i \neq j \in \epsilon} k_{ij}^{z_{s3}} (f_{ij} \cdot l_{ij}) \iota \otimes \partial_t |r_j|. \quad (5.8)$$

Here  $l_{ij}$  is a dimensionless vector defined as

$$l_{ij} = \left( r_{ij} \cdot \frac{r_{ij}}{\|r_{ij}\|} \right) \frac{r_{ij}}{\|r_{ij}\|}, \quad (5.9)$$

where  $r_{ij}$  is the position vector of material point  $j$  with respect to  $i$  (note that  $r_{ij} = -r_{ji}$ );  $k_{ij}^{z_{s3}}$  is a dimensionless quantity defined by

$$k_{ij}^{z_{s3}} = \frac{\|r_{ij}\| - \|r_{ij}^0\| \|r_j\|}{\|r_{ij}\| - \|r_{ij}^0\| r_{gir}^0}, \quad (5.10)$$

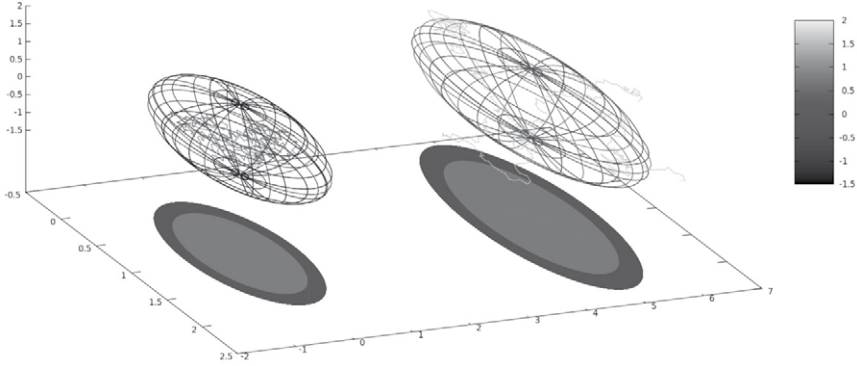


Figure 5.25: Comparison of the ellipsoids (and associated proteins) for the two cases already compared in panels *a*, *b* and *c* of Figure 5.24. It is apparent which is the molecule unfolding just from the difference in the shape of the ellipsoids, which refer to the square root of the reciprocals of the eigenvalues of  $U$ . Snapshots are taken when the difference between the first eigenvalues is close to the maximum one, according to Figure 5.24*a*. The palette indicates the  $z$ -axis values of the upper half portion of the ellipsoids. Protein dimensions are normalized by  $r_{gir}^0$ .

where,  $r_{ij}^0$  is the placement vector of residue  $j$  referred to residue  $i$  in the reference configuration of the molecule and other symbols have been already defined. Finally  $\iota$  is a vector the entries of which are all unity.

In this approach the first quantity involved in the tensor product is an effective measure of the local action between amino acids  $i$  and  $j$ , while the second quantity, once summed upon all points and integrated in time, is a gross measure of the evolution of the distances within the molecule thought as a whole. Since the actions practically result from the projection of the forces exchanged between amino acids  $i$  and  $j$  onto the vector joining them, only peptide interactions and native contacts (Lennard-Jones potentials) are considered in the computation of  $z_{s3}$ . The essential idea behind this approach is that  $\partial_t |r_j|$  governs the evolution of  $z_{s3}$ . Therefore the whole left side term of the tensorial product, *i.e.*  $k_{ij}^{z_{s3}} (f_{ij} \cdot l_{ij}) \iota$  is tailored in a way that it has a constant sign when the forces are generated by harmonic or harmonic-like potentials, as peptide bonds or Lennard-Jones interactions, respectively. Here ‘harmonic-like potentials’ means potentials that are attractive in a region of the phase space and repulsive in the complementary one. For a matter of clarity we shall refer to this approach as SAM3 in the present section.

An alternative formulation that stems from this model and actually reduces the whole evolution of  $\nu$  to the evolution of its first diagonal entry, is subsequently described and will be addressed to as SAM3\*. This approach allows one to drop the absolute value of the quantity  $r_j$  in the time partial derivative of (5.8),

without losing the ability to discerning between detachment or approach of an amino-acid with respect to the center of mass. The aim is accomplished by simply evaluating the vector  $r_j$  in a local frame of reference that has its  $x$ -axis coincident with the vector itself, namely each residue has a local proper frame of reference where its placement vector (with respect to the center of mass) is evaluated. In the specific case, the placement vector itself fixes the  $x$ -axis of such a coordinate system. Thence only the first entry of  $r_j$  is different from zero and so only the first diagonal component of  $z_{s3}$  evolves. In practice, the order parameter is reduced from a second-rank tensor to a scalar. However, tensor rank can be easily restored claiming a volume-preserving deformation: in this way the other two diagonal entries of  $\nu$  can be deduced by requiring that the ellipsoid associated with them would have a constant volume, equal to the initial one (the volume of the material element). Since there is no preferred assumption, in the volume-preserving enforcement the two indeterminate axes (*i.e.* the ones that are not governed by the self-action) are set equal to each other. Similarly to the second approach introduced in section 5.1, both the square root of the reciprocals of the eigenvalues and  $\nu$  diagonal entries are considered for SAM3, while for SAM3\* only the latter ones are evaluated.

**Main results.** In this section all results for SAM3 and SAM3\* are briefly outlined. Generally speaking, the formulations analyzed here require the computation of several additional quantities or extra frames for every time-step during a simulation, resulting more time-consuming than the two previous ones. For this reason, no translocation simulations have been performed. Moreover, the already presented outcomes are deemed appropriate for the purpose of the present investigation. However, a few results are pointed out as the quality of the description for both SAM3 and SAM3\* is adequate. As far as (5.8) is concerned, the diagonal components of  $\nu$  provide a clear description of the MBP shape in all the runs performed, as it was expected. In this case, the value of  $\epsilon$  can be very small and so  $\nu$  entries very large. As it results from Figure. 5.26,  $\nu(1, 1)$  leads the dynamics and closely follows the trend of the associated bounding box edge in the central principal frame for an AFM-like stretching simulation. The evolution of the remaining two entries is negligible, although panel *b* and *c* highlight the fact that their trends are not so close to the edges of the limiting parallelepiped. Usually the appropriateness of the picture improves in the PDB system of coordinates.

We find that only two of the eigenvalues of  $U$  evolve, as one of them is constantly equal to 1 regardless the involved dynamics. Hence the picture associated with the  $\lambda_i^{-1/2}$  is just two-dimensional, as show in Figure 5.26*d* and relative inset. This 2D description is in a sense in accord with the evolution of the gross shape of the molecule (as basically one eigenvalue forms a peak and the other one a minimum), but was not expected.

As already stated, it is in the PDB frame of reference that SAM3 works better. As it is depicted in Figure 5.27,  $\nu$  diagonal entries closely follow the evolution of the bounding box edges during a thermalization simulation. The agreement is considerable and the curves reported in both panels *a* and *b* are hardly distinguishable (this holds true also for the diagonal component not depicted). Quantitative agreement is just a matter of fine tuning the material element volume.

SAM3\* approach generates a 1D evolution of the self-action and therefore also of the associated  $\nu$ . However, the 3D evolution can be recovered artificially claiming a volume-preserving transformation for the ellipsoid, as already mentioned. In such a way it is probably obtained the most appropriate description of all the ones presented in the thesis, as the constant-volume assumption allows both  $\lambda_2^{-1/2}$  and  $\lambda_3^{-1/2}$  to considerably decrease from the initial value. The ellipsoid picture is drafted in Figure 5.28 for a stretching simulation, considering both the trends of the square root of the reciprocals of the eigenvalues (upper picture) and a snapshot of the system close to the maximum elongation of the molecule (panel *b*). From the latter panel it is immediately appreciable the quality of the description as the almost single file conformation of the MBP gives rise to an ellipsoid that is an extremely thin and long spindle.

### 5.3. Remarks

Three additional formulations of  $z_s$  have been considered to overcome the limitations of the basic approach. Among them, the first one (sec. 5.1) shows the best premises, as it is computationally inexpensive and allows a proper description of the molecule shape in all the simulations performed. SAM3 and SAM3\* approaches require the computation of several additional quantities at every time step, considerably slowing down the efficiency of the numerical code. Also, the symmetric matrix for SAM3 results two-dimensional, regardless the simulated dynamics. SAM3\*, when associated to a volume-preserving deformation requirement, generates an evolution of the ellipsoid in close agreement with the real deformation of MBP. However, no translocation simulations have been performed for SAM3 and SAM3\*. Also, systems smaller than the MBP have not been studied in this frameworks. However, there are no doubts that the appropriateness of the approaches here analyzed will not be compromised by the dimensionality of the system, since the associated  $z_s$  formulations arise from considerations on deformation mechanisms between two single material points.

What is necessary to further develop are the physical consequences of selecting empirical formulations for  $z_s$ , as the description of the material element shape is only the first step to be undertaken in the building up of a continuum model for a complex body.

## Coarse-grained molecular dynamics and continuum models

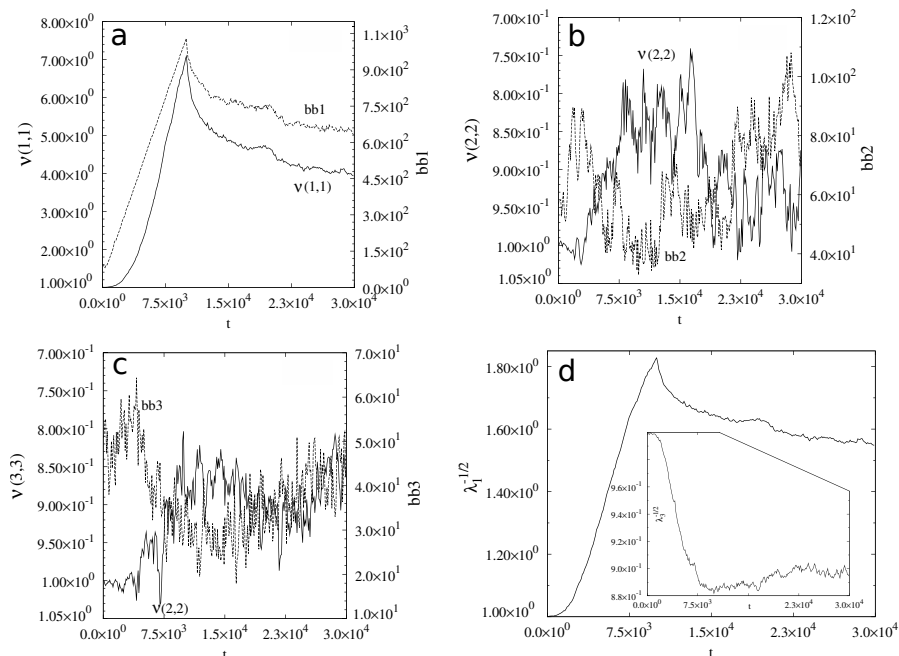


Figure 5.26: Mechanical stretching simulation for MBP. Panel *a*: Comparison between  $\nu(1, 1)$  and the bounding box  $x$ -axis edge in the central principal frame,  $bb_1$ .  $\nu$  entry ends up leading the whole dynamics of the ellipsoid and it shows a trend totally in compliance with the gross shape evolution of the MBP. Panel *b* and *c*:  $\nu(2, 2)$  and  $\nu(3, 3)$  components compared with the associated edges. The trends are not so in agreement as in panel *a*, but it is appropriate to point out that amplitudes are both negligible if compared with  $\nu(1, 1)$ . Panel *d*: This panel and the relative inset shows the trends of the two evolving symmetric matrix eigenvalues. The third eigenvalue (not shown) is always constant, independently from the performed simulation. Hence the ellipsoid evolution associated to  $U$  is only 2D for SAM3. However  $\lambda_1^{-1/2}$  correctly reproduces the elongation while  $\lambda_3^{-1/2}$  depicts the shrinking of the shape. We note that symmetric matrix eigenvalues undergo oscillations that are considerably smaller than the ones of the diagonal entries of  $\nu$ .

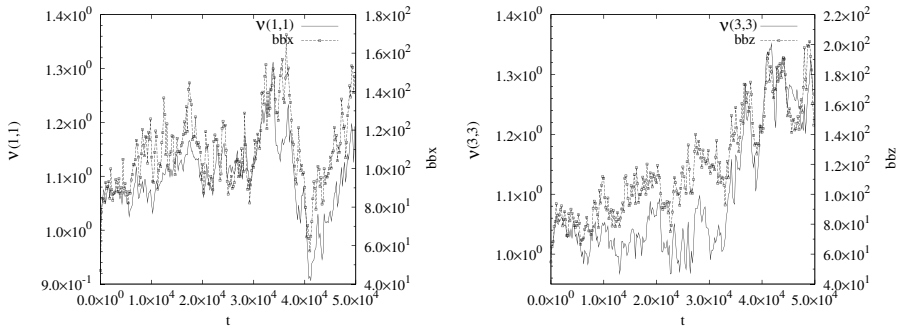


Figure 5.27: Panel *a*: Here the trends of  $\nu(1, 1)$  and the  $x$ -axis edge of the bounding box in the PDB frame are compared. Trends are so close to each other that a visual distinction is tough. It is in the PDB frame of reference that SAM3 achieves the best description of the MBP shape as showed in this case for the challenging thermalization case. Panel *b*: Similarly to panel *a*,  $\nu(3, 3)$  and  $bb_z$  are compared, pointing out again the close similarity.  $\nu(2, 2)$  and  $bb_y$  trends are not shown but their agreement is comparable with the ones reported in this figure.

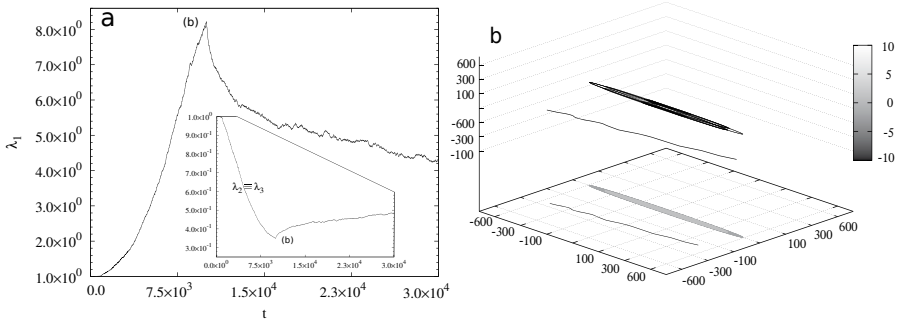


Figure 5.28: Panel *a*: Trends of the square root of the reciprocals of the eigenvalues associated to SAM3\* for a mechanical stretching simulation of MBP.  $\lambda_2^{-1/2}$  and  $\lambda_3^{-1/2}$  are obtained claiming a constant volume deformation for the ellipsoid and are set equal to each other. The trends result in close agreement with the overall shape deformation of the protein in a AFM-like stretching simulation. Panel *b*: Visual comparison between the almost linear shape of the MBP at maximum elongation and the associated ellipsoid, which has been obtained from the data of panel *a* at maximum stretching multiplied by the radius of gyration of the native MBP (also, a magnification coefficient of 2 for the semi axes has been necessary, although a fine tuning of  $\epsilon$  would have worked the same). The affinity of the two shapes is here substantial as the constant volume assumption allows to achieve an adequate shrinking of both the transverse axes of the ellipsoid.





## References

- [1] N. Gō and H.A. Scheraga. *On the use of classical statistical mechanics in the treatment of polymer chain conformations*. «Macromolecules», 9(4):535–542, 1976.
- [2] C. Clementi, H. Nymeyer, and J.N. Onuchic. *Topological and energetic factors: what determines the structural details of the transition state ensemble and “on-route” intermediates for protein folding? An investigation for small globular proteins*. «J. Mol. Biol.», 298:937–953, 2000.
- [3] P.M. Mariano. *Multifield theories in the mechanics of solids*. «Adv. Appl. Mech.», 38:1–93, 2002.
- [4] W. Wickner and R. Schekman. *Protein translocation across biological membranes*. «Science», 310(5753):1452–1456, 2005.
- [5] J.O. Bustamante, J.A. Hanover, and A. Liepins. *The ion channel behavior of the nuclear pore complex*. «J. Membr. Biol.», 146(3):239–251, 1995.
- [6] B. Hanss, E. Leal-Pinto, L.A. Bruggeman, T.D. Copeland, and P.E. Klotman. *Identification and characterization of a cell membrane nucleic acid channel*. «Proc. Natl. Acad. Sci. USA», 95(4):1921–1926, 1998.
- [7] M.S. Sanford and Blobel G. *A protein-conducting channel in the endoplasmic reticulum*. «Cell», 65(3):371–380, 1991.
- [8] A. Matouschek. *Protein unfolding - an important process in vivo?* «Curr. Opin. Struct. Bio.», 13(1):98–109, 2003.
- [9] S. Prakash and A. Matouschek. *Protein unfolding in the cell*. «Trends Biochem Sci.», 29(11):593–600, 2004.
- [10] J.J. Kasianowicz, E. Brandin, D. Branton, and D.W. Deamer. *Characterization of individual polynucleotide molecules using a membrane channel*. «Proc. Natl. Acad. Sci. USA», 93(24):13770–13773, 1996.
- [11] D.E. Makarov. *Computer simulations and theory of protein translocation*. «Acc. Chem. Res.», 42(2):281–289, 2008.
- [12] A. Matouschek and C. Bustamante. *Finding a protein’s achilles heel*. «Nat. Struct. Biol.», 10(9):674, 2003.
- [13] R.W. DeBlois and C.P. Bean. *Counting and sizing of submicron particles by the resistive pulse technique*. «Rev. Sci. Instrum.», 41(7):909–916, 1970.
- [14] W. Humphrey, A. Dalke, and K. Schulten. *VMD: visual molecular dynamics*. «J. Mol. Graphics Modell.», 14(1):33–38, 1996.
- [15] O. P. Hamill, A. Marty, E. Neher, B. Sakmann, and F. J. Sigworth. *Improved patch-clamp techniques for high-resolution current recording from cells and*

- cell-free membrane patches*. «Pflügers Archiv European Journal of Physiology», 391(2):85–100, 1981.
- [16] A. Meller, L. Nivon, E. Brandin, J. Golovchenko, and D. Branton. *Rapid nanopore discrimination between single polynucleotide molecules*. «Proc. Natl. Acad. Sci. USA», 97(3):1079–1084, 2000.
- [17] O. Braha, B. Walker, S. Cheley, J.J. Kasianowicz, L. Song, J.E. Gouaux, and H. Bayley. *Designed protein pores as components for biosensors*. «Chemistry & biology», 4(7):497–505, 1997.
- [18] J.W.F. Robertson, C.G. Rodrigues, V.M. Stanford, K.A. Rubinson, O.V. Krasilnikov, and J.J. Kasianowicz. *Single-molecule mass spectrometry in solution using a solitary nanopore*. «Proc. Natl. Acad. Sci. USA», 104(20):8207–8211, 2007.
- [19] D. Pedone, M. Firnkes, and U. Rant. *Data analysis of translocation events in nanopore experiments*. «Anal. Chem.», 81(23):9689–9694, 2009.
- [20] J.J. Kasianowicz, E. Brandin, M. Akesson, D. Branton and D.W. Deamer. *Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single rna molecules*. «Biophys. J.», 77(6):3227–3233, 1999.
- [21] A. Meller, L. Nivon, and D. Branton. *Voltage-driven dna translocations through a nanopore*. «Phys. Rev. Lett.», 86(15):3435–3438, 2001.
- [22] M. Bates, M. Burns, and A. Meller. *Dynamics of dna molecules in a membrane channel probed by active control techniques*. «Biophys. J.», 84:2366–2372, 2003.
- [23] G.A. Oukhaled, J. Mathe, A.L. Biance, L. Bacri, J.M. Betton, D. Lairez, J. Pelta, and L. Auvray. *Unfolding of proteins and long transient conformations detected by single nanopore recording*. «Phys. Rev. Lett.», 98(15):15810, 2007.
- [24] L. Movileanu. *Interrogating single proteins through nanopores: challenges and opportunities*. «Trends Biotechnol.», 27(6):333–341, 2009.
- [25] E. Gouaux.  *$\alpha$ -hemolysin from staphylococcus aureus: An archetype of  $\beta$ -barrel, channel-forming toxins*. «J. Struct. Biol.», 121(2):110–122, 1998.
- [26] M.W. Parker, J.T. Buckley, J.P.M. Postma, A.D. Tucker, K. Leonard, and F. Pattus. *Structure of the aeromonas toxin proaerolysin in its water-soluble and membrane-channel states*. «Nature», 367(6460):292–295.
- [27] M. Pastoriza-Gallego, L. Rabah, G. Gibrat, B. Thiebot, F.G. van der Goot, L. Auvray, J.M. Betton, and J. Pelta. *Dynamics of unfolded protein transport through an aerolysin pore*. «J. Am. Chem. Soc.», 133(9):2923–2931, 2011.
- [28] C. Merstorf, B. Cressiot, M. Pastoriza-Gallego, G.A. Oukhaled, J.M. Betton, L. Auvray, and J. Pelta. *Wild type, mutant protein unfolding and phase transition detected by single-nanopore recording*. «ACS Chem. Biol.», 7(4):652–658, 2012.
- [29] S. Howorka, L. Movileanu, X. Lu, M. Magnon, S. Cheley, O. Braha, and H. Bayley. *A protein pore with a single polymer chain tethered within the*

- lumen*. «J. Am. Chem. Soc.», 122(11):2411–2416, 2000.
- [30] L. Movileanu, S. Howorka, O. Braha, and H. Bayley. *Detecting protein analytes that modulate transmembrane movement of a polymer chain within a single protein pore*. «Nature Biotechnology», 18(10):1091–1095, 2000.
- [31] C.Y. Kong and M. Muthukumar. *Simulations of stochastic sensing of proteins*. «J. Am. Chem. Soc.», 127(51):18252–18261, 2005.
- [32] T.C. Sutherland, Y.T. Long, R.I. Stefureac, I. Bediako-Amoa, H.B. Kraatz, and J.S. Lee. *Structure of peptides investigated by nanopore analysis*. «Nano Letters», 4(7):1273–1277, 2004.
- [33] L. Movileanu, J.P. Schmittschmitt, J.M. Scholtz, and H. Bayley. *Interactions of peptides with a protein pore*. «Biophys. J.», 89(2):1030–1045, 2005.
- [34] R. Stefureac, Y.T. Long, H.B. Kraatz, P. Howard, and J.S. Lee. *Transport of  $\alpha$ -helical peptides through  $\alpha$ -hemolysin and aerolysin pores*. «Biochemistry», 45(30):9172–9179, 2006.
- [35] C.P. Goodrich, S. Kirmizialtin, B.M. Huyghues-Despointes, A. Zhu, J.M. Scholtz, D.E. Makarov, and L. Movileanu. *Single-molecule electrophoresis of  $\beta$ -hairpin peptides by electrical recordings and langevin dynamics simulations*. «J. Phys. Chem. B», 111(13):3332–3335, 2007.
- [36] S. Zhang, E. Udho, Z. Wu, R.J. Collier, and A. Finkelstein. *Protein translocation through anthrax toxin channels formed in planar lipid bilayers*. «Biophys. J.», 87(6):3842–3849, 2004.
- [37] M. Pastoriza-Gallego, G. Gibrat, B. Thiebot, J.M. Betton, and J. Pelta. *Polyelectrolyte and unfolded protein pore entrance depends on the pore geometry*. «Biochim. Biophys. Acta - Biomembranes», 1788(6):1377–1386, 2009.
- [38] D.S. Talaga and J. Li. *Single-molecule protein unfolding in solid state nanopores*. «J. Am. Chem. Soc.», 131(26):9287–9297, 2009.
- [39] C.M. Niemeyer and Mirkin A.C. *Nanobiotechnology: Concepts, Applications and Perspectives*. Wiley-VHC Verlag, 2005.
- [40] L.Q. Gu, O. Braha, S. Conlan, S. Cheley, and H. Bayley. *Stochastic sensing of organic analytes by a pore-forming protein containing a molecular adapter*. «Nature», 398(6729):686–690, 1999.
- [41] A. F. Sauer-Budge, J. A. Nyamwanda, D. K. Lubensky, and D. Branton. *Unzipping kinetics of double-stranded dna in a nanopore*. «Phys. Rev. Lett.», 90(23):238101, Jun 2003.
- [42] J. Mathé, H. Visram, V. Viasnoff, Y. Rabin, and A. Meller. *Nanopore unzipping of individual dna hairpin molecules*. «Biophys. J.», 87(5):3205–3212, 2004.
- [43] B.A. Krantz, A. Finkelstein, and R.J. Collier. *Protein translocation through the anthrax toxin transmembrane pore is driven by a proton gradient*. «J. Mol. Biol.», 355(5):968–979, 2006.
- [44] C. Dekker. *Solid state nanopores*. «Nature Nanotech.», 2(4):209–215, 2007.
- [45] K. Healy, B. Schiedt, and A.P. Morrison. *Solid-state nanopore technologies for nanopore-based dna analysis*. «Nanomedicines», 2(6):875–897, 2007.

- [46] E. Pennisi. *Search for pore-fection*. «Science», 336(6081):534–537, 2012.
- [47] Z. Siwy, L. Trofin, P. Kohli, L.A. Baker, C. Trautmann, and C.R. Martin. *Protein biosensors based on biofunctionalized conical gold nanotubes*. «J. Am. Chem. Soc.», 127(14):5000–5001, 2005.
- [48] D. Fologea, M. Gershow, B. Ledden, D.S. McNabb, J.A. Golovchenko, and J. Li. *Detecting single stranded dna with a solid state nanopore*. «Nano Letters», 5(10):1905–1909, 2005.
- [49] A. Han, G. Schurmann, G. Mondin, R.A. Bitterli, N.G. Hegelbach, N.F. de Rooij, and U. Staufer. *Sensing protein molecules using nanofabricated pores*. «Appl. Phys. Lett.», 88(9):093901, 2006.
- [50] D. Fologea, B. Ledden, D.S. McNabb, and J. Li. *Electrical characterization of protein molecules by a solid-state nanopore*. «Appl. Phys. Lett.», 91(5):053901, 2007.
- [51] G.A. Oukhaled, B. Cressiot, L. Bacri, M. Pastoriza-Gallego, J.M. Betton, E. Bourhis, R. Jede, J. Gierak, L. Auvray, and J. Pelta. *Dynamics of completely unfolded and native proteins through solid-state nanopores as a function of electric driving force*. «ACS Nano», 5(5):3628–3638, 2011.
- [52] B. Cressiot, G.A. Oukhaled, G. Patriarche, M. Pastoriza-Gallego, J.M. Betton, L. Auvray, M. Muthukumar, L. Bacri, and J. Pelta. *Protein transport through a narrow solid-state nanopore at high voltage: Experiments and theory*. «ACS nano», 6(7):6236–6243, 2012.
- [53] D. Branton and *et al.* *The potential and challenge of nanopore sequencing*. «Nature Biotechnology», 26(10):1146–1153, 2008.
- [54] L. Huang, S. Kirmizialtin, and D.E. Makarov. *Computer simulations of the translocation and unfolding of a protein pulled mechanically through a pore*. «J. Chem. Phys.», 123:124903, 2005.
- [55] A. Ammenti, F. Cecconi, U. Marini-Bettolo-Marconi, and A. Vulpiani. *A statistical model for translocation of structured polypeptide chains through nanopores*. «J. Phys. Chem. B», 113(30):10348–10356, 2009.
- [56] M. Bacci, M. Chinappi, C.M. Casciola, and F. Cecconi. *Role of denaturation in maltose binding protein translocation dynamics*. «J. Phys. Chem. B», 116(14):4255–4262, 2012.
- [57] S. Redner. *A guide to first-passage processes*. Cambridge University Press, Cambridge, first edition, 2001.
- [58] J.L. Ericksen. *Kinematics of macromolecules*. «Arch. Rational Mech. Anal.», 9(1):1–8, 1962.
- [59] P.M. Mariano. *Slow motion in granular matter*. «Lect. Notes Math.», 1937:193–210, 2008.
- [60] M. Balsera, S. Stepaniants, S Izrailev, Y Oononi, and Schulten K. *Reconstructing potential energy functions from simulated force-induced unbinding processes*. «Biophys. J.», 73(3):1281–1287, 1997.
- [61] P.C. Li and D. E Makarov. *Theoretical studies of the mechanical unfolding of the muscle protein titin: Bridging the time-scale gap between simulation*

- and experiment.* «J. Chem. Phys.», 119(17):9260–9268, 2003.
- [62] S. Lichter, B. Rafferty, Z. Flohr, and A. Martini. *Protein high-force pulling simulations yield low-force results.* «PLoS ONE», 7(4):e34781, 04 2012.
- [63] D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan, and W. Wriggers. *Atomic-level characterization of the structural dynamics of proteins.* «Science», 330(6002):341–346, 2010.
- [64] P.L. Freddolino, C.B. Harrison, Y. Liu, and K. Schulten. *Challenges in protein-folding simulations.* «Nat. Phys.», 6(10):751–758.
- [65] K. Lindorff-Larsen, S. Piana, R.O. Dror, and D.E. Shaw. *How fast-folding proteins fold.* «Science», 334(6055):517–520, 2011.
- [66] L. Huang and D.E. Makarov. *Translocation of a knotted polypeptide through a pore.* «J. Chem. Phys.», 129:121107, 2008.
- [67] A.G. Oukhaled, A.L. Biance, J. Pelta, L. Auvray, and L. Bacri. *Transport of long neutral polymers in the semidilute regime through a protein nanopore.* «Phys. Rev. Lett.», 108(8):88104, 2012.
- [68] D. Rotem, L. Jayasinghe, M. Salichou, and H. Bayley. *Protein detection by nanopores equipped with aptamers.* «J. Am. Chem. Soc.», 134(5):2781–2787, 2012.
- [69] C. Madampage, O. Tavassoly, C. Christensen, M. Kumari, and J.S. Lee. *Nanopore analysis: An emerging technique for studying the folding and misfolding of proteins.* «Prion», 6(2):110–118, 2012.
- [70] R. Stefureac, L. Waldner, P. Howard, and J.S. Lee. *Nanopore analysis of a small 86-residue protein.* «Small», 4(1):59–63, 2008.
- [71] D. Fologea, B. Ledden, D.S. McNabb, and J. Li. *Electrical characterization of protein molecules by a solid-state nanopore.* «Appl. Phys. Lett.», 91(5):053901, 2007.
- [72] G.A. Oukhaled, L. Bacri, M. Pastoriza-Gallego, J.M. Betton, and J. Pelta. *Sensing proteins through nanopores: Fundamental to applications.* «ACS Chem. Biol.», 2012.
- [73] T.A. Desai, D.J. Hansford, L. Kulinsky, A.H. Nashat, G. Rasi, J. Tu, Y. Wang, M. Zhang, and M. Ferrari. *Nanopore technology for biomedical applications.* «Biomed. Microdevices», 2(1):11–40, 1999.
- [74] W. Sung and P.J. Park. *Polymer translocation through a pore in a membrane.* «Phys. Rev. Lett.», 77(4):783–786, 1996.
- [75] D.K. Lubensky and D.R. Nelson. *Driven polymer translocation through a narrow pore.* «Biophys. J.», 77(4):1824–1838, 1999.
- [76] A.M. Berezhkovskii, M.A. Pustovoit, and S.M. Bezrukov. *Channel-facilitated membrane transport: Transit probability and interaction with the channel.* «J. Chem. Phys.», 116:9952–9956, 2002.
- [77] A.M. Berezhkovskii and I.V. Gopich. *Translocation of rod-like polymers through membrane channels.* «Biophys. J.», 84:787–793, 2003.

- [78] R. Metzler and J. Klafter. *When translocation dynamics becomes anomalous*. «Biophys. J.», 85(4):27762779, 2003.
- [79] E. Slonkina and A.B. Kolomeisky. *Polymer translocation through a long nanopore*. «J. Chem. Phys.», 118:7112–7118, 2003.
- [80] A. Aksimentiev, J.B. Heng, G. Timp, and K. Schulten. *Microscopic kinetics of dna translocation through synthetic nanopores*. «Biophys. J.», 87(3):2086–2097, 2004.
- [81] M. Muthukumar and C.Y. Kong. *Simulation of polymer translocation through protein channels*. «Proc. Natl. Acad. Sci. USA», 103(14):5273–5278, 2006.
- [82] J. Mathé, A. Aksimentiev, D. R. Nelson, K. Schulten, and A. Meller. *Orientation discrimination of single-stranded dna inside the  $\alpha$ -hemolysin membrane channel*. «Proc. Natl. Acad. Sci. USA», 102(35):12377–12382, 2005.
- [83] S. Matysiak, A. Montesi, M. Pasquali, A.B. Kolomeisky, and C. Clementi. *Dynamics of Polymer Translocation through Nanopores: Theory Meets Experiment*. «Phys. Rev. Lett.», 96(11):118103, 2006.
- [84] P. Tian and I. Andricioaei. *Size, motion, and function of the secy translocon revealed by molecular dynamics simulations with virtual probes*. «Biophys. J.», 90(8):2718–2730, 2006.
- [85] J. Gumbart and K. Schulten. *Molecular dynamics studies of the archaeal translocon*. «Biophys. J.», 90(7):2356–2367, 2006.
- [86] P. Tian and I. Andricioaei. *Repetitive pulling catalyzes co-translocational unfolding of barnase during import through a mitochondrial pore*. «J. Mol. Biol.», 350(5):1017–1034, 2005.
- [87] A.H. de Vries S.J. Marrink and A.E. Mark. *Coarse grained model for semi-quantitative lipid simulations*. «J. Phys. Chem. B», 108(2):750–760, 2004.
- [88] S.T. Yefimov P.D. Tieleman S.J. Marrink, H.J. Risselada and A.H. de Vries. *The MARTINI force field: Coarse grained model for biomolecular simulations*. «J. Phys. Chem. B», 111(27):7812–7824, 2007.
- [89] C. Guardiani, R. Livi, and F. Cecconi. *Coarse grained modeling and approaches to protein folding*. «Curr. Bioinf.», 5(3):217–240, 2010.
- [90] J.M. Sorenson and T. Head-Gordon. *Matching simulation and experiment: a new simplified model for simulating protein folding*. «J. Comp. Bio.», 7(3-4):469–481, 2000.
- [91] J.D. Honeycutt and D. Thirumalai. *The nature of folded states of globular proteins*. «Biopolymers», 32(6):695–709, 1992.
- [92] S. Brown, N.J. Fawzi, and T. Head-Gordon. *Coarse-grained sequences for protein folding and design*. «Proc. Natl. Acad. Sci. USA», 100(1):10712–10717, 2003.
- [93] E.I. Shakhnovich. *Protein design: a perspective from simple tractable models*. «Fold. Des.», 3(3):R45–R58, 1998.
- [94] R. Livi F. Cecconi, C. Guardiani. *Testing simplified protein models of the hpin1 ww domain*. «Biophys. J.», 91(2):694–704, 2006.

- [95] G.M. Torrie and J.P. Valleau. *Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling*. «J. Comput. Phys.», 23(2):187–199, 1977.
- [96] M. Souaille and B. Roux. *Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations*. «Comput. Phys. Commun.», 135(1):40–57, 2001.
- [97] F.A. Quirocho, J.C. Spurlino, and L.E. Rodseth. *Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor*. «Structure», 5(8):997–1015, 1997.
- [98] M. Bertz and M. Rief. *Mechanical unfoldons as building blocks of maltose-binding protein*. «J. Mol. Biol.», 378(2):447–458, 2008.
- [99] C. Hyeon and D. Thirumalai. *Capturing the essence of folding and functions of biomolecules using coarse-grained models*. «Nat. Commun.», 2:2:487, 2011.
- [100] N. Gō. *Theoretical studies of protein folding*. «Ann. Rev. of Biophys. and Bioeng.», 12(1):183, 1983.
- [101] E. Shakhnovich. *Theoretical studies of protein-folding thermodynamics and kinetics*. «Curr. Opin. Struct. Bio.», 7(1):29–40, 1997.
- [102] D. Baker. *A surprising simplicity to protein folding*. «Nature», 405(6782):39–42, 2000.
- [103] J.K. Karanicolas and C.L. Brooks. *Improved Gō-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions*. «J. Mol. Bio.», 334(2):309–325, 2003.
- [104] J.N. Onuchic and P.G. Wolynes. *Theory of protein unfolding*. «Curr. Opin. Struct. Bio.», 14(1):70–75, 2004.
- [105] R.D. Jr. Hills and C.L. Brooks. *Insights from coarse-grained Gō models for protein folding and dynamics*. «Int. J. of Mol. Sci.», 10(3):889–905, 2009.
- [106] W. Kabsch. *A solution for the best rotation to relate two sets of vectors*. «Acta Crystallogr. Sect. A», 32(5):922–923, 1976.
- [107] S. Takada. *Gō-ing for the prediction of protein folding mechanisms*. «Proc. Natl. Acad. Sci. USA», 96:11698–11700, 1999.
- [108] M. Cieplak and T.X. Hoang. *Sequencing of folding events in Gō-like proteins*. «J. Chem. Phys.», 113:8319–8328, 2000.
- [109] M. Cieplak, T.X. Hoang, and M.O. Robbins. *Folding and stretching in a Gō-like model of tiltin*. «Proteins: Struct., Func., Gent.», 49(1):114–124, 2002.
- [110] H. Kaya and H.S. Chan. *Contact order dependent protein folding rates: Kinetic consequences of a cooperative interplay between favorable nonlocal interactions and local conformational preferences*. «Proteins: Struct., Funct., Genet.», 52(4):524–533, 2003.
- [111] M. Knott, H. Kaya, and H.S. Chan. *Energetics of protein thermodynamic cooperativity: contributions of local and nonlocal interaction*. «Polymer», 45(2):623–632, 2008.



- [112] F. Cecconi, Guardiani C., and Livi R. *Stability and kinetic properties of c5-domain for myosin binding protein c and its mutants.* «Biophys. J.», 94(4):1403–1411, 2008.
- [113] F. Cecconi, Guardiani C., and Livi R. *Computational analysis of folding and mutation properties of c5 domain from myosin binding protein c.* «Proteins: Struct., Funct., Bioinf.», 70(4):1313–1322, 2008.
- [114] M. Chinappi, F. Cecconi, and C.M. Casciola. *Computational analysis of maltose binding protein translocation.* «Philos. Mag.», 91(13-15):2034–2048, 2011.
- [115] O.K. Dudko, T.G.W. Graham, and R.B. Best. *Locating the barrier for folding of single molecules under an external force.* «Phys. Rev. Lett.», 107(20):208301, 2011.
- [116] L. Verlet. *Computer "experiments" on classical fluids. I. Thermodynamical properties of lennard-jones molecules.* «Phys. Rev.», 159(1):98–103, Jul 1967.
- [117] H. Lu and K. Schulten. *Steered molecular dynamics simulation of conformational changes of immunoglobulin domain i27 interpret atomic force microscopy observations.* «J. Chem. Phys.», 247(1):141–153, 1999.
- [118] C. Ganesh, A.N. Shah, C.P. Swaminathan, A. Surolia, and R. Varadarajan. *Thermodynamic characterization of the reversible, two-state unfolding of maltose binding protein, a large two-domain protein.* «Biochemistry», 36(16):5020–5028, 1997.
- [119] J.F. Marko and E.D. Siggia. *Stretching dna.* «Macromolecules», 28(26):8759–8770, 1995.
- [120] C. Clementi. *Coarse-grained models of protein folding: toy models or predictive tools?* «Curr. Opin. Struct. Bio.», 18(1):10–15, 2008.
- [121] J.J. Kasianowicz and *et al.* *Driven dna transport into an asymmetric nanometer-scale pore.* «Phys. Rev. Lett.», 85(14):3057–33060, 2000.
- [122] R.B. Best and G. Hummer. *Reaction coordinates and rates from transition paths.* «Proc. Natl. Acad. Sci. USA», 102(19):6732–6737, 2005.
- [123] M. Muthukumar. *Polymer translocation through a hole.* «J. Chem. Phys.», 111(22):10371–10374, 1999.
- [124] D. Bouzida S.H. Robert S. Kumar, J.M. Rosenberg and P.A. Kollman. *The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method.* «J. Comput. Chem.», 13(8):1011–1021, 1992.
- [125] G.J. Davies, M. Dauter, A.M. Brzozowski, M.E. Bjornvad, K.V. Andersen, and M. Schulein. *Structure of the bacillus agaradherans family 5 endoglucanase at 1.6 Å and its cellobiose complex at 2.0 Å resolution.* «Biochemistry», 37(7):1926–1932, 1998.
- [126] M. Muthukumar. *Translocation of a confined polymer through a hole.* «Phys. Rev. Lett.», 86(14):3188–3191, Apr 2001.
- [127] J. K. Wolterink, G.T. Barkema, and D. Panja. *Passage times for unbiased polymer translocation through a narrow pore.* «Phys. Rev. Lett.», 96(20):208301, May 2006.

- [128] D.E. Makarov S. Kirmizialtin, L. Huang. *Topography of free-energy landscape via mechanical unfolding of proteins*. «J. Chem. Phys.», 122:234915–234927, 2005.
- [129] S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten. *Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality*. «J. Chem. Phys.», 119:3559–3566, 2003.
- [130] J.L. Ericksen. *On the cauchy-born rule*. «Math. & Mech. of Solids», 13(3-4):199–220, 2008.
- [131] W. E. *Principles of multiscale modeling*. Cambridge University Press, Cambridge, 2011.
- [132] P.M. Mariano and F.L. Stazi. *Computational aspects of the mechanics of complex materials*. «Arch. Comp. Meth. E», 12(4):391–487, 2005.



## PREMIO TESI DI DOTTORATO

### ANNO 2007

- Bracardi M., *La Materia e lo Spirito. Mario Ridolfi nel paesaggio umbro*  
Coppi E., *Purines as Transmitter Molecules. Electrophysiological Studies on Purinergic Signalling in Different Cell Systems*  
Mannini M., *Molecular Magnetic Materials on Solid Surfaces*  
Natali I., *The Ur-Portrait. Stephen Hero ed il processo di creazione artistica in A Portrait of the Artist as a Young Man*  
Petretto L., *Imprenditore ed Università nello start-up di impresa. Ruoli e relazioni critiche*

### ANNO 2008

- Bemporad F., *Folding and Aggregation Studies in the Acylphosphatase-Like Family*  
Buono A., *Esercito, istituzioni, territorio. Alloggiamenti militari e «case Herme» nello Stato di Milano (secoli XVI e XVII)*  
Castenasi S., *La finanza di progetto tra interesse pubblico e interessi privati*  
Colica G., *Use of Microorganisms in the Removal of Pollutants from the Wastewater*  
Gabbiani C., *Proteins as Possible Targets for Antitumor Metal Complexes: Biophysical Studies of their Interactions*

### ANNO 2009

- Decorosi F., *Studio di ceppi batterici per il biorisanamento di suoli contaminati da Cr(VI)*  
Di Carlo P., *I Kalasha del Hindu Kush: ricerche linguistiche e antropologiche*  
Di Patti F., *Finite-Size Effects in Stochastic Models of Population Dynamics: Applications to Biomedicine and Biology*  
Inzitari M., *Determinants of Mobility Disability in Older Adults: Evidence from Population-Based Epidemiologic Studies*  
Macrì F., *Verso un nuovo diritto penale sessuale. Diritto vivente, diritto comparato e prospettive di riforma della disciplina dei reati sessuali in Italia*  
Pace R., *Identità e diritti delle donne. Per una cittadinanza di genere nella formazione*  
Vignolini S., *Sub-Wavelength Probing and Modification of Complex Photonic Structures*

### ANNO 2010

- Fedi M., *«Tuo lumine». L'accademia dei Risvegliati e lo spettacolo a Pistoia tra Sei e Settecento*  
Fondi M., *Bioinformatics of genome evolution: from ancestral to modern metabolism. Phylogenomics and comparative genomics to understand microbial evolution*  
Marino E., *An Integrated Nonlinear Wind-Waves Model for Offshore Wind Turbines*  
Orsi V., *Crisi e Rigenerazione nella valle dell'Alto Khabur (Siria). La produzione ceramica nel passaggio dal Bronzo Antico al Bronzo Medio*  
Polito C., *Molecular imaging in Parkinson's disease*  
Romano R., *Smart Skin Envelope. Integrazione architettonica di tecnologie dinamiche e innovative per il risparmio energetico*

### ANNO 2011

- Acciaioli S., *Il trompe-l'œil letterario, ovvero il sorriso ironico nell'opera di Wilhelm Hauff*  
Bernacchioni C., *Sfingolipidi bioattivi e loro ruolo nell'azione biologica di fattori di crescita e citochine*  
Fabbri N., *Bragg spectroscopy of quantum gases: Exploring physics in one dimension*  
Gordillo Hervás R., *La construcción religiosa de la Hélade imperial: El Panhelenion*  
Mugelli C., *Indipendenza e professionalità del giudice in Cina*  
Pollastri S., *Il ruolo di TAF12B e UVR3 nel ciclo circadiano dei vegetali*  
Salizzoni E., *Paesaggi Protetti. Laboratori di sperimentazione per il paesaggio costiero euro-mediterraneo*

ANNO 2012

- Adessi A., *Hydrogen production using Purple Non-Sulfur Bacteria (PNSB) cultivated under natural or artificial light conditions with synthetic or fermentation derived substrates*
- Alberti M., *La 'scoperta' dei disoccupati. Alle origini dell'indagine statistica sulla disoccupazione nell'Italia liberale (1893-1915)*
- Bondi D., *Filosofia e storiografia nel dibattito anglo-americano sulla svolta linguistica*
- Evangelisti E., *Structural and functional aspects of membranes: the involvement of lipid rafts in Alzheimer's disease pathogenesis. The interplay between protein oligomers and plasma membrane physicochemical features in determining cytotoxicity*
- Gualdani R., *Using the Patch-Clamp technique to shed light on ion channels structure, function and pharmacology*
- Petrucci F., *Petri Candidi Decembrii Epistolarum iuveniliium libri octo. A cura di Federico Petrucci*
- Ramalli A., *Development of novel ultrasound techniques for imaging and elastography. From simulation to real-time implementation*

ANNO 2013

- Bacci M., *Coarse-grained molecular dynamics and continuum models for the transport of protein molecules*
- Brancasi I., *Architettura e Illuminismo. Filosofia e progetti di città nel tardo Settecento francese*
- Cucinotta E., *Produzione poetica e storia nella prassi e nella teoria greca di età classica*
- Locatelli M., *Mid infrared digital holography and terahertz imaging*
- Lunghi C., *Early cross-modal interactions and adult human visual cortical plasticity revealed by binocular rivalry*
- Martelli R., *Characteristics of raw and cooked fillets in species of actual and potential interest for Italian aquaculture. Rainbow trout (*Oncorhynchus mykiss*) and meagre (*Argyrosomus regius*)*
- Muniz Miranda F., *Modelling of spectroscopic and structural properties using molecular dynamics*
- Pellegrini L., *Circostanze del reato: trasformazioni in atto e prospettive di riforma*



