

**TIME-RESOLVED  
FORCE DISTRIBUTION ANALYSIS  
FOR MOLECULAR COMMUNICATION**

Dissertation  
submitted to the  
Combined Faculties for the Natural Sciences and for Mathematics  
of the Ruperto-Carola University of Heidelberg, Germany  
for the degree of  
Doctor of Natural Sciences

presented by  
Diplom-Biochemiker Ion Bogdan Costescu  
born in Bucharest, Romania



Dissertation  
submitted to the  
Combined Faculties for the Natural Sciences and for Mathematics  
of the Ruperto-Carola University of Heidelberg, Germany  
for the degree of  
Doctor of Natural Sciences

presented by  
Diplom-Biochemiker Ion Bogdan Costescu  
born in Bucharest, Romania



**TIME-RESOLVED  
FORCE DISTRIBUTION ANALYSIS  
FOR MOLECULAR COMMUNICATION**

REFEREES:

Prof. Dr. Robert B. Russel

Dr. Frauke Gräter



*To my family.*

*I am among those who think that science has great beauty.  
A scientist in his laboratory is not only a technician: he is also a child  
placed before natural phenomena which impress him like a fairy tale.*

*Marie Curie (1867 - 1934)*

---

# ABSTRACT

---

During communication processes, such as the allosteric regulation of enzymatic activity or the opening of mechanically-gated channels, biomolecules undergo conformational changes. Perturbations, like the binding of a ligand or stress in the cell membrane, act similar to the application of an external force. They trigger a response which is directly dependent on the receptor's mechanical properties, which can be characterized, among others, through Atomic Force Microscopy (AFM) experiments.

We introduce here Time-Resolved Force Distribution Analysis (TRFDA), an extension to Molecular Dynamics (MD) simulations, providing an engineer's view to *in silico* (bio)molecules. TRFDA reveals the internal distribution of forces and stresses in equilibrium states or during conformational changes, thereby exposing the propagation of an external force throughout the molecular structure. We also introduce the concept of punctual stress, an expression of stress at atomic level, which highlights the structural elements involved in the mechanical response.

We first apply TRFDA to study the indentation until rupture of a single layer graphene sheet, providing essential insight into the distribution of stress in molecular structures during material deformation and rupture. We find that stress accumulates under the AFM indenter tip much stronger than previously assumed<sup>1,2</sup>, and it decays to almost background levels at distances as low as 5–10 times the indenter radius. The graphene rupture is initiated by thermal fluctuations in the stressed material, and the probability of rupture decreases exponentially as the distance from the indenter tip increases, explaining the locality of material failure observed experimentally<sup>1</sup>.

The force-induced unfolding of two small proteins, ubiquitin and NuG2, represents a second application of TRFDA. In force-clamp MD simulations, we reproduce the stretched exponential kinetics reported experimentally<sup>3,4</sup> for the unfolding of these proteins. The unfolding kinetic curves become more stretched with the decreasing mechanical resistance of the protein, suggesting that the two-state kinetic model of protein unfolding should be augmented by a component expressing the protein elasticity. This finding is in line with the theory of glassy dynamics<sup>3</sup> or static disorder in the transition states<sup>5</sup>, proposed earlier as explanation for the stretched exponential kinetics. For different applied forces, we

determine unfolding rates, then compare and combine them with experimental kinetic data in a single model<sup>6,7</sup>, predicting equilibrium kinetic parameters which agree remarkably well with experimental ones. Using TRFDA, we identify the structural elements bearing most of the external force, and find that similarity in the secondary and tertiary structure is not a good predictor of similarity of unfolding mechanisms and mechanical properties. Our analysis of internal forces and stresses in tensed proteins also suggests that the stretched exponential kinetics is simply an expression of the protein elasticity.

By providing a dynamic view of forces and stress variations in MD simulations of dynamic processes, TRFDA can give insights in the structure and functionality of biomolecules, and is the ideal tool to complement experimental techniques in determining mechanical properties. We therefore hope that TRFDA will soon become a common tool for analyzing results from MD simulations of (bio)materials.

---

# ZUSAMMENFASSUNG

---

Bei Kommunikationsprozessen wie der allosterischen Regulation von Enzymen oder der Öffnung von mechanisch gesteuerten Kanälen durchlaufen Biomoleküle Konformationsänderungen. Störungen, wie die Bindung eines Liganden oder Stress in der Zellmembran, wirken ähnlich wie die Anwendung einer externen Kraft. Diese Störungen lösen eine Reaktion aus, welche direkt von den mechanischen Eigenschaften des Rezeptors abhängt, welche wiederum mit einem Rasterkraftmikroskop (AFM) charakterisiert werden können.

In dieser Arbeit führen wir eine zeitaufgelöste Kraftverteilungsanalyse (TRFDA) als eine Erweiterung von Molekulardynamik-Simulationen ein, welche einen Einblick in (Bio-)Moleküle aus der Perspektive eines Ingenieurs gibt. Mit Hilfe TRFDA lässt sich die interne Kraft- und Stressverteilung im Gleichgewichtszustand oder während Konformationsänderungen aufzeigen. Dabei wird die Propagation der externen Kraft durch die molekulare Struktur offengelegt. Wir führen auch das Konzept eines punktuellen Stresses ein, ein Ausdruck mechanischen Stresses auf der atomistischen Größenskala. Dadurch ist es uns möglich, strukturelle Elemente zu identifizieren, welche in der mechanischen Reaktion auf die äußere Störung involviert sind.

Als erstes wenden wir TRFDA an, um die Verbiegung einer Monolage Graphen bis zum Zerreißen zu studieren. Dadurch erhalten wir einen essentiellen Einblick in die Stressverteilung innerhalb des molekularen Systems während der Materialdeformation bis zum Bruch. Wir stellen fest, dass der Stress unter der Spitze des Kraftmikroskops viel stärker akkumuliert als bisher angenommen<sup>1,2</sup> und abfällt schon bei dem 5–10 fachen Abstand des Spitzenradius bis fast auf Hintergrundwerte. Der Riss von Graphenlagen wird von thermischen Fluktuationen des unter Stress stehenden Materials herbeigeführt, wobei die Risswahrscheinlichkeit der chemischen Bindungen in Graphen exponentiell mit zunehmenden Abstand zur Spitze abnimmt. Das erklärt die Lokalität des Materialversagens, die experimentell beobachtet wurde<sup>1</sup>.

Die kraftinduzierte Entfaltung von zwei kleinen Proteinen, Ubiquitin und NuG2, repräsentiert eine zweite Anwendung von TRFDA. In MD Simulationen unter konstanter Kraft reproduzieren wir die gleiche - unerwarteterweise gestreckte - exponentielle Kinetik wie

beim experimentellen Entfalten von Proteinen beobachtet<sup>3,4</sup>. Die kinetischen Entfaltungskurven werden mit abfallendem mechanischen Widerstand mehr gestreckt, was ein kinetisches Modell mit zwei Zuständen suggeriert, das um eine Komponente zur Proteinelastizität erweitert werden sollte. Diese Beobachtung steht im Einklang mit Theorien zu “glassy dynamics”<sup>3</sup> oder auch zur “static disorder in the transition states”<sup>5</sup>. Beide Theorien wurden in vorangegangenen Arbeiten zur Erklärung der gestreckten exponentiellen Kinetik vorgeschlagen. Wir bestimmen und vergleichen Entfaltungsraten für verschiedene angewandte Kräfte, fassen sie mit experimentellen kinetischen Daten in einem einzelnen Modell zusammen<sup>6,7</sup> und berechnen daraus kinetische Gleichgewichtsparameter, welche ausserordentlich gut mit experimentellen Daten übereinstimmen. Mittels TRFDA identifizieren wir strukturelle Elemente, welche den grössten Anteil externer Kraft aufnehmen, und stellen fest, dass eine Ähnlichkeit in der sekundären und tertiären Struktur eines Proteins kein gutes Vorhersagekriterium für die Ähnlichkeit im Entfaltungsmechanismus oder in den mechanischen Eigenschaften darstellt. Unsere Analyse von internen Kräften und Stress in vorgespannten Proteinen suggeriert auch, dass die gestreckte exponentielle Kinetik in erster Linie ein Ausdruck für die Proteinelastizität darstellt.

Durch Bereitstellen eines dynamischen Abbilds der Kräfte- und Stressvariationen in MD Simulationen kann TRFDA Einblicke in die Struktur und Funktionalität von Biomolekülen geben und ist ein ideales Werkzeug, um Experimente zur Charakterisierung mechanischer Eigenschaften zu komplementieren. Wir hoffen deshalb, dass TRFDA bald ein weit verbreitetes Werkzeug für die Analyse von MD Simulationen von (Bio-)Materialien wird.

---

# PUBLICATIONS

---

In connection with this work, the following articles have been published or prepared for submission into peer reviewed international journals:

1. B.I. Costescu and F. Gräter - Time-Resolved Force Distribution Analysis. *BMC Biophysics* (2013). *In press*.
2. B.I. Costescu, I.B. Baldus and F. Gräter - First Principles Based Morse Potential for Efficient Molecular Dynamics Simulations of Graphene. *Submitted*.
3. B.I. Costescu and F. Gräter - Atomic stress distribution during indentation of graphene until rupture. *Submitted*.
4. B.I. Costescu and F. Gräter - Role of mechanical resistance in forced unfolding of proteins. *In preparation*.



---

# ACKNOWLEDGMENTS

---

Many people have encouraged and supported me during my work on this thesis.

First and foremost, I would like to express my deep gratitude to Dr. Frauke Gräter for offering me the opportunity to work on this topic, for her confidence in my abilities, for being always open for discussions, and for guiding me through many challenging tasks while also giving me the freedom to follow my own (sometimes too original) ideas. I am also indebted to Prof. Dr. Robert Russell for taking over the role of the primary supervisor, for his patience, the constructive discussions and very useful suggestions, which considerably improved this work.

I was lucky to have as deskmates Christian, Cedric and Ilona, with whom I shared most of the pleasure and pain of being a PhD student. Agnieszka and Camilo made available their postdoctoral wisdom and experience through discussions that offered new insights in my research subjects. I'd like to express many thanks to the entire Molecular Biomechanics group for making research exciting and fun, and for all the after-hours distractions from science. Extended thanks to all the HITSters who made the institute a nice working place.

I am grateful to Prof. Dr. Peter Bastian for taking over the role of my supervisor in the frame of the HGS MathComp, for his patience and support. The diligence and dedication of Dr. Kirsten Hei turned the Biosciences Kenntnisstandprfung into an instructive and enjoyable experience.

Special thanks to my Romanian friends in Heidelberg who have been the source of much relaxation and joy over the years. Their company has been essential in fueling my good spirits.

Most importantly, I would like to thank my family for their constant and unconditional support, encouragement and understanding, for their unlimited patience with my erratic schedule in the past three years, and for always being a solid shoulder to lean on.



---

# CONTENTS

---

1	INTRODUCTION	1
1.1	Atomic Force Microscopy . . . . .	2
1.2	The Bell model . . . . .	4
1.3	Overview . . . . .	6
2	TIME-RESOLVED FORCE DISTRIBUTION ANALYSIS	9
2.1	Molecular Dynamics simulations . . . . .	9
2.1.1	Interaction potentials . . . . .	10
2.1.2	Integration . . . . .	12
2.1.3	Short range and long range interactions . . . . .	13
2.1.4	Periodic boundary conditions . . . . .	14
2.1.5	Temperature and pressure control . . . . .	15
2.1.6	Applying an external force . . . . .	17
2.2	Time-Resolved Force Distribution Analysis . . . . .	18
2.2.1	Decomposition of forces from 3- and 4-body potentials . . . . .	19
2.2.2	Internally computed pairwise forces between residues . . . . .	20
2.2.3	Vector to scalar transformations . . . . .	21
2.2.4	Noise reduction . . . . .	21
2.3	Stress at atomic level in MD simulations . . . . .	22
2.3.1	Atomic virial stress . . . . .	23
2.3.2	Local pressure . . . . .	23
2.3.3	Punctual stress . . . . .	24
2.3.4	Circular stress . . . . .	25
2.3.5	Comparison of atomic stress definitions . . . . .	26
3	GRAPHENE RUPTURE	29
3.1	Introduction . . . . .	29
3.2	A truncated Morse potential . . . . .	32
3.3	Graphene indentation until rupture . . . . .	35
3.4	Discussion and conclusions . . . . .	45

## CONTENTS

4	MECHANICAL UNFOLDING OF PROTEINS	49
4.1	Introduction . . . . .	49
4.2	Structural aspects of protein unfolding . . . . .	51
4.3	Force-clamp MD erases molecular memory . . . . .	56
4.4	Non-exponential unfolding kinetics . . . . .	58
4.5	Kinetic parameters in the absence of force . . . . .	64
4.6	Internal distribution of forces and stresses . . . . .	68
4.6.1	Distribution and propagation of mechanical stress . . . . .	68
4.6.2	Forces in hydrogen bonds . . . . .	71
4.6.3	Force distribution in stretched proteins . . . . .	74
4.7	Summary and conclusions . . . . .	81
5	CONCLUSIONS AND FURTHER PERSPECTIVES	85
A	MULTIBODY FORCE DECOMPOSITION	91
A.1	Force decomposition for 3-body potentials . . . . .	91
A.2	Force decomposition for 4-body potentials . . . . .	92
B	GRAPHENE RUPTURE - METHODS	95
C	PROTEIN UNFOLDING - METHODS	99
	BIBLIOGRAPHY	103

---

# LIST OF FIGURES

---

Figure 2.1	The concept of pairwise forces . . . . .	18
Figure 2.2	Noise in pairwise forces . . . . .	21
Figure 2.3	Noise in punctual stress . . . . .	25
Figure 3.1	Carbon atoms in a single layer graphene sheet . . . . .	30
Figure 3.2	The AFM indentation of a graphene sheet . . . . .	31
Figure 3.3	The modified Morse potential . . . . .	32
Figure 3.4	Graphene patches used for the truncated Morse potential parametriza- tion . . . . .	33
Figure 3.5	Phases of the indentation process until rupture . . . . .	35
Figure 3.6	Force-indentation profiles . . . . .	36
Figure 3.7	Variability of rupture force, and the influence of sphere velocity, sphere radius and graphene sheet radius. . . . .	37
Figure 3.8	Two-dimensional elastic modulus . . . . .	38
Figure 3.9	Punctual stress variation during the indentation process until rup- ture . . . . .	39
Figure 3.10	Punctual, virial and circular stress during indentation . . . . .	40
Figure 3.11	Punctual stress distribution just before rupture . . . . .	41
Figure 3.12	Circular and virial stress distribution just before rupture . . . . .	42
Figure 3.13	Contact area between tip and graphene sheet . . . . .	43
Figure 3.14	Probability and localization of bond breaking . . . . .	44
Figure 4.1	Structure of ubiquitin and NuG2 . . . . .	51
Figure 4.2	Typical evolution of the end-to-end distance during forced protein unfolding . . . . .	53
Figure 4.3	Evolution of the end-to-end distance from 500 forced unfolding simulations of ubiquitin and NuG2 . . . . .	54
Figure 4.4	End-to-end distance histograms . . . . .	55
Figure 4.5	Correlation of unfolding times . . . . .	57
Figure 4.6	Kinetic profiles from forced unfolding of ubiquitin . . . . .	59
Figure 4.7	Bell and DHS model fit for ubiquitin . . . . .	65

## List of Figures

Figure 4.8	Bell and DHS model fit for NuG2 . . . . .	67
Figure 4.9	Stress during equilibrium and immediately after application of force	69
Figure 4.10	Stress during the overall waiting time . . . . .	70
Figure 4.11	Pairwise forces in H-bonds between the $\beta_1$ and $\beta_5$ strands of ubiquitin . . . . .	72
Figure 4.12	Pairwise forces in H-bonds between the $\beta_1$ and $\beta_4$ strands of NuG2	73
Figure 4.13	Pairwise forces as a function of the applied force . . . . .	74
Figure 4.14	Pairwise forces in ubiquitin . . . . .	75
Figure 4.15	Pairwise forces in NuG2 . . . . .	77
Figure 4.16	NuG2 unfolding steps . . . . .	79
Figure 4.17	Frequency of occurrence of the two unfolding mechanisms for NuG2	80
Figure 4.18	Kinetic profiles of the two unfolding mechanisms of NuG2 . . . . .	81
Figure A.1	Force decomposition for an angle potential . . . . .	92
Figure A.2	Force decomposition for a dihedral angle potential . . . . .	93

---

# LIST OF TABLES

---

Table 4.1	Frequency of occurrence of higher end-to-end distance plateaus . . . . .	55
Table 4.2	Stretched exponential parameters from ubiquitin unfolding . . . . .	62
Table 4.3	Stretched exponential parameters from NuG2 unfolding . . . . .	62
Table 4.4	Equilibrium kinetic parameters for ubiquitin . . . . .	66
Table 4.5	Equilibrium kinetic parameters for ubiquitin . . . . .	66
Table 4.6	Probability of existence of H-bonds between the $\beta_1$ and $\beta_5$ strands of ubiquitin . . . . .	71
Table 4.7	Probability of existence of H-bonds between the $\beta_1$ and $\beta_4$ strands of NuG2 . . . . .	72
Table 4.8	Projections of pairwise forces on the end-to-end distance in ubiquitin	76
Table 4.9	Projections of pairwise forces on the end-to-end distance in NuG2	78



---

# ACRONYMS

---

AFM - Atomic Force Microscopy

CDF - cumulative distribution function

DHS - Dudko, Hummer and Szabo model

FDA - Force Distribution Analysis

FE - Finite Elements

MD - Molecular Dynamics

PBC - Periodic Boundary Conditions

PME - Particle Mesh Ewald

TRFDA - Time-Resolved Force Distribution Analysis



---

# INTRODUCTION

---

In biological systems, molecules are very rarely alone. Cells, as well as the inter-cellular environment close to the cell surface, are very crowded spaces, in which molecules come into contact very often with other molecules. During this contact, molecules can exert forces on one another - as is usually the case in ligand binding leading to allosteric regulation of biological processes, or can transfer small molecular fragments or electrical charges.

Such processes are the basis for communication, an essential feature of living organisms, at all levels. A eukaryotic cell dies if left alone, due to the missing communication with other cells or its environment. Different types of tissues associated into an organ need to communicate to perform the required function. Different organs of a mammal exchange information, for example through hormones, cooperating to allow the organism to survive. Metabolic or cell signaling networks form the underlying mechanism for most of these communication processes.

At the cellular level, communication is the basis for tissue homeostasis, repair and development, or for immunity. Errors in cellular communication pathways can lead to cell death or malfunctioning, as encountered in diseases like cancer or autoimmune response. Cells can communicate with each other through a direct contact, or over short or long dis-

tances, each with specific signal transmission mechanisms. Mechanical signal transduction is often encountered across membranes, allowing, for example, communications between the nucleus and the rest of the cell or between the cell and the extracellular environment, including other cells. Mechanical signaling is often involved in protein allostery, in which the binding of a ligand or effector at a site leads to regulation of the protein activity through conformational changes of the active site<sup>8</sup>. Mechanical signaling can also occur through stretch-gated ion channels, which respond to stress in the cell membrane by opening or closing, generating electrical currents, changes in pH, or transport of small molecules<sup>8</sup>.

The mechanosensors receive the signal in form of an external force. Although the ligand binding has long been considered to follow the lock and key model, it is currently believed that the allosteric site can undergo local conformational changes in response to the presence of the ligand<sup>8</sup>, similar to mechanical deformations caused by an external force. The membrane tension acts directly by stretching of ion channels. Thus, studies of mechanical deformation of proteins can be straightforwardly related to such communication mechanisms<sup>9</sup>.

The recent development of experimental techniques, like Atomic Force Microscopy, optical tweezers or biomembrane force probes<sup>10</sup>, has allowed studying in unprecedented detail the mechanical response of biomolecules to an external force. These techniques can be used to apply or measure mechanical forces at a single molecule level, providing details related to the structure, stability and activity of proteins, nucleic acids or membranes. Forced unfolding of proteins has demonstrated the intimate relationship between their molecular geometry and mechanical properties<sup>9</sup>, and therefore constitutes an important method of study in this work. Furthermore, the mechanical response has been shown to depend on the orientation of the force<sup>11</sup>, as well as on its magnitude<sup>6,7,12</sup>. In the rest of this chapter, we briefly present the Atomic Force Microscopy method and its extension, force spectroscopy, which is nowadays one of the preferred ways to study forced unfolding of proteins, as well as the theoretical background for dependency of the protein mechanical response, in form of unfolding rates, on the applied force.

## 1.1 ATOMIC FORCE MICROSCOPY

Atomic Force Microscopy (AFM) is a technique initially developed for imaging non-conductive surfaces, like polymers, ceramics, glass, or biological samples. The principle is simple: a hard tip is moved on the horizontal surface and the vertical displacement of the tip is recorded, generating a nano-level representation of the geometry of the surface. Most often, the tip has a rounded shape, is made of Si or Si<sub>3</sub>N<sub>4</sub>, has a radius of several nm to several tens of nm, and is attached to a planar cantilever whose displacement is measured

through a laser beam deflection system. The scanning can cover areas of up to a hundredth of  $\text{mm}^2$  in several minutes and is able to reach true atomic resolution. Although the large scanning time might lead to degradation of biological samples, the high precision makes it a precious tool in the laboratory.

Apart from surface imaging, the atomic force microscope can also be used to apply a force to a surface or sample, in a technique called *force spectroscopy*. The force is not measured directly, but calculated from the deflection of the cantilever of known stiffness through Hooke's law. The tip can be moved towards the sample or away from it. In the first case, it presses onto the sample, generating an indentation. In the second case, it leads to a stretching of the sample, which is fixed on a substrate. The tip plays a special role in force spectroscopy experiments. When indenting a sample, the tip has to resist the applied force; this is straightforward for biological samples which are soft, but can become a difficult problem when the material of study is strong, like ceramics or graphene (Chapter 3). In this mode, forces up to several microNewtons can be reached<sup>1</sup>. When the sample is stretched, it adheres to the tip through weak attractive van der Waals, and possibly electrostatic, interactions; therefore, the sample can easily detach from the tip, such that reaching high stretching forces or keeping the sample stretched for long times is difficult or even impossible. In this mode, the measuring resolution is limited to around 10 pN by the thermal noise.

When applying a force to a sample, the atomic force microscope can function in two modes: constant force and constant velocity. To keep a constant force, a feedback mechanism is used to adjust the cantilever position such that the force is kept at the desired value; the result of the experiment is the variation in time of the cantilever position. When the tip is moved with constant velocity, the variations in force are recorded and the results are often represented as force versus indentation curves. Both these modes have applications in biology; while constant velocity experiments are used to measure rupture forces, for example of ligand-protein complexes, the constant force studies are preferred to describe protein kinetics.

In this work, we refer to, and attempt to reproduce as well as possible, experiments of force spectroscopy leading to graphene rupture<sup>1</sup> (Chapter 3) and protein unfolding<sup>3,11,13,14</sup> (Chapter 4). Such studies have already provided interesting details related to the protein structure-function relationship, such as the functioning as mechanosensors, import and degradation in proteasome, or the structural basis of the underlying mechanical resistance<sup>9</sup>.

## 1.2 THE BELL MODEL

Many biological processes involve the association between two or more molecules, such as enzymes and substrates, or antigens and antibodies, mediated by forming and breaking of interactions. Analogously, interactions within a single molecule can break upon the transition between conformational states, such as the folded and unfolded state of a protein, as considered in this work (Chapter 4). From an energetic point of view, these processes are assumed to have two states - bound and unbound, or folded and unfolded, separated by a barrier which needs to be crossed for the process to happen. The transition between states is a temperature-dependent process, as expressed by Arrhenius' law:

$$k_0 = Ae^{-\Delta G/k_B T} \quad (1.1)$$

where  $k_0$  is the reaction rate,  $A$  is a reaction-specific constant,  $\Delta G$  is the free-energy barrier,  $k_B$  is the Boltzmann constant, and  $T$  is the temperature. Bell introduced the idea that not only temperature but also mechanical force can have an influence on the reaction rate<sup>6</sup>. An external constant force  $F$ , applied during the transition between states at constant temperature, performs work equal to  $F\Delta x$ , where  $\Delta x$  is the distance along the reaction coordinate between the reactant and the transition state. This work lowers the energy barrier, such that the reaction rate becomes force-dependent, following:

$$k_0 = Ae^{-(\Delta G - F\Delta x)/k_B T} \quad (1.2)$$

The change in rate by force can then be expressed as

$$k_F = k_0 e^{F\Delta x/k_B T} \quad (1.3)$$

where  $k_F$  is the rate constant at force  $F$ ; of note here is the absence of the reaction-specific constant  $A$  and free-energy barrier  $\Delta G$ . By taking the natural logarithm, the relation becomes

$$\ln k_F = \ln k_0 + \frac{F\Delta x}{k_B T} \quad (1.4)$$

The Bell model makes several assumptions. First, it assumes that the transition along the reaction coordinate  $x$  is independent of any other (possibly force-dependent) process; this implies that the free-energy profile remains one-dimensional, with the reactant and product states separated by a single energy barrier. Second, it assumes that the reaction kinetics continues to be described by an exponential law at a force  $F > 0$  just as it was at

$F = 0$ . Finally, it considers the distance between the reactant and transition state  $\Delta x$  to be independent from  $F$ , such that the energy barrier is lowered but not shifted by the external force.

To account for cases where the latter assumption is not valid, Dudko, Hummer and Szabo have proposed another model<sup>7,12</sup> (DHS) which also includes the effect of the change in the free-energy landscape upon force application, resulting in an overall decreasing  $\Delta x$ . The DHS model also contains the apparent free-energy of activation  $\Delta G$  as an additional parameter and can be expressed in terms of reaction rates as:

$$k_F = k_0 \alpha^{(1/\nu)-1} e^{\Delta G / (k_B T) (1-\alpha^{1/\nu})} \quad (1.5)$$

where  $\alpha = 1 - \nu F \Delta x / \Delta G$ , and  $\nu$  is a factor expressing the nature of the underlying free-energy profile for the transition between states.  $\nu = 1/2$  corresponds to a harmonic well with a cusp-like barrier,  $\nu = 2/3$  represents a potential containing a combination of linear and cubic terms, and  $\nu = 1$  recovers Bell's model. The DHS formalism predicts that both the logarithm of the bound state lifetime under constant force conditions and the mean breaking force are nonlinear functions of the force and the logarithm of the rate of force application, respectively. The model was derived for forces at which a free-energy barrier of several  $k_B T$  still exists, which implies that the external force has to be kept below a critical force at which the barrier vanishes  $F < \Delta G / (\nu \Delta x)$ .

The Bell model has been successfully applied to kinetic data obtained from experiments or from MD simulations. However, the MD derived results differ significantly from the experimental ones, and this has been hypothesized to be the consequence of different ranges of forces and velocities that can be reached in each type of study. The DHS formalism tries to bridge this gap, offering a unified model spanning several orders of magnitude. Still, we can continue to assume the Bell model as valid for small ranges of forces, where the free-energy landscape can be approximated by a linear function of the external force.

Although introduced to express the force dependency of bond breaking rates, the Bell model can be more generally used for any process involving a transition between two states under the influence of an external force, for which the rate is assumed to follow an exponential law. In this work, the Bell model will be applied to the rupture of graphene upon indentation, to represent both the probability of C-C bond rupture as function of the bond force and the sheet rupture force dependency on the indentation velocity (Chapter 3). It will also be applied to protein unfolding under mechanical force, to express the relationship between the unfolding rate and the applied force (Chapter 4).

## 1.3 OVERVIEW

With Molecular Dynamics (MD) simulations, we attempt to reproduce as close as possible force spectroscopy experiments of the mechanical indentation of a single graphene sheet<sup>1</sup> or of protein unfolding<sup>3,14</sup>, in order to study the distribution of forces and stresses throughout the molecules and to relate it to mechanical properties or function.

In Chapter 2, we briefly present the simulation and analysis techniques used throughout this work. The application of external forces is a central part of this study; Section 2.1.6 explains how mechanical forces can be exerted during MD simulations and gives some examples of using them in biologically relevant studies. We then introduce the Time-Resolved Force Distribution Analysis (TRFDA), a method developed within this work, which allows the study of distribution and propagation of forces and stresses at atomic level, from MD simulations. As the concept of stress at atomic level is not well defined, we propose two new definitions based on the concept of pairwise forces, as calculated from TRFDA.

Chapter 3 presents the application of TRFDA on a (mostly) two-dimensional molecular system. TRFDA is here used to calculate pairwise forces and stresses in a single-atom thick graphene sheet, during mechanical deformation until rupture. In order to represent material rupture, we first derive a truncated Morse potential which allows representing the carbon-carbon bond breaking. We then validate our model of the graphene sheet by comparing it to experimental and theoretical data. Finally, we apply TRFDA to reveal the distribution of punctual and circular stress in the sheet and compare them with the virial atomic stress. TRFDA also allows us to link the probability and location of material rupture to the bond breaking. As part of this study, we learn about the variability of punctual stress and pairwise forces, and how effective averaging pairwise forces in equivalent bonds is in reducing noise from thermal fluctuations.

The mechanical unfolding of proteins is the subject of Chapter 4. Here, we study two proteins (ubiquitin and NuG2) which are very similar in structure, yet have been shown to have a different unfolding kinetic behavior<sup>3,14</sup>. To characterize the unfolding process, we consider both the unfolding times, as in the force spectroscopy experiments, and the distribution of forces or stresses in the three-dimensional structure of the proteins, as obtained from TRFDA. We reproduce the stretched exponential kinetics<sup>3</sup> for a large range of forces applied on both proteins. We then calculate zero force kinetic parameters from the Bell and DHS models (Section 1.2), and attempt to investigate their validity for the ranges of forces used in experiments and MD simulations. We then turn to use TRFDA, in order to characterize the stress distribution in the stretched proteins, to elucidate their mechanism of forced unfolding, and to suggest an explanation for the stretched exponential kinetics.

Finally, in Chapter 5, we summarize the important results of these studies, present some of the challenges we encountered during our work, and propose some new directions which attempt to improve and extend TRFDA, or use it to elucidate further mysteries of mechanical signaling or resistance of proteins.



# 2

---

## TIME-RESOLVED FORCE DISTRIBUTION ANALYSIS

---

This chapter describes the simulation and analysis techniques used throughout this work. It starts with a presentation of the main aspects of Molecular Dynamics (MD) simulations. It then introduces the Time-Resolved Force Distribution Analysis (TRFDA), a method which we developed for studying the distribution and propagation of forces and stresses at atomic level. As the concept of stress at atomic level is not well defined, the last part of the chapter proposes two alternative definitions, which use pairwise forces obtained from TRFDA, and compares them with the existing definitions.

### 2.1 MOLECULAR DYNAMICS SIMULATIONS

The Molecular Dynamics (MD) simulations represent a technique of calculating the time dependent behavior of systems composed of one or more molecules, each molecule containing one or more atoms. Atoms are represented as spheres and interact through spring-like potentials, forming a classical many-body system. The state of the molecular system is

described by the combination of atom positions in the three-dimensional space and of instantaneous atom velocities. The progress in time from one state to the next is achieved through integration of Newton's equations of motion.

The method was initially used in the late 1950s to study the interactions of a few hundreds hard spheres<sup>15,16</sup>. Several years later, the simulation of almost 1000 atoms of liquid argon used for the first time a realistic (Lennard-Jones) potential<sup>17</sup>. As the method continued to be developed, more complex molecules could be studied - of biological interest are the first simulations of liquid water<sup>18</sup> and of a protein (bovine pancreatic trypsin inhibitor)<sup>19</sup>, both of which were performed in the 1970s. Since then, improvements in computational power, development of force fields and methodological advances have made MD simulations a well-established tool to study all kinds of biological molecules: proteins, nucleic acids, lipids and, more recently, sugars, as well as complexes thereof. Nowadays, MD simulations have a wide range of applications. For example, they are routinely used to perform ligand binding studies as part of the development of new drugs<sup>20</sup>, can be combined with quantum mechanical (QM) calculations to simulate bond formation or breaking in enzymatic reactions<sup>21,22</sup>, and have been used for *in silico* folding or unfolding of small proteins<sup>23,24</sup>.

### 2.1.1 Interaction potentials

MD simulations can provide the details of individual atomic motions as a function of time. They can also be used to describe system-wide physical properties of a molecular model, in many cases even more easily than experiments performed on the real system. To achieve such performance, they require a representation of the energy of the system in form of a force field. This combines an empirical set of functions and their respective parameters, defining the energy due to interactions between the different atomic species present in the simulated system as a function of the positions of the atoms. A typical force field used for biomolecules, like OPLS-AA<sup>25</sup>, is composed of bonded and non-bonded terms. Bonded terms (bonds, angles, dihedral angles and improper dihedral angles) represent intra-molecular interactions between atoms forming chemical bonds; non-bonded terms (van der Waals, electrostatic) describe intra- or inter-molecular interactions between atoms which might be some distance apart. The terms present in the OPLS-AA force field, which will be used throughout this work, express the energy of interaction between atoms through the following functional forms:

$V_{\text{bond}}$  is the interaction energy due to the bond formed by atoms  $i$  and  $j$ ,

$$V_{\text{bond}}(r_{ij}) = \frac{1}{2}k_{ij}^b(r_{ij} - b_{ij})^2 \quad (2.1)$$

with  $r_{ij}$  being the distance between atoms  $i$  and  $j$ , and  $k_{ij}^b$  and  $b_{ij}$  being the force constant and equilibrium distance for the bond.

$V_{\text{angle}}$  is the interaction energy due to the angle  $\theta_{ijk}$  formed by the bonds between atoms  $i$ ,  $j$  and  $j$ ,  $k$ ,

$$V_{\text{angle}}(\theta_{ijk}) = \frac{1}{2} k_{ijk}^{\theta} (\theta_{ijk} - \theta_{ijk}^0)^2 \quad (2.2)$$

with  $k_{ijk}^{\theta}$  and  $\theta_{ijk}^0$  being the force constant and equilibrium value for the angle.

$$V_{\text{dihedral}}(\phi_{ijkl}) = \sum_{n=0}^5 C_n (\cos(\phi_{ijkl} - 180^\circ))^n \quad (2.3)$$

is the Ryckaert-Bellemans dihedral angle potential energy function, where  $C_n$  are coefficients and  $\phi_{ijkl}$  is the dihedral angle formed by the two planes defined by atoms  $i$ ,  $j$ ,  $k$  and  $j$ ,  $k$ ,  $l$ .

$V_{\text{improper}}$  is the potential energy of the  $\phi_{ijkl}$  dihedral angle,

$$V_{\text{improper}}(\phi_{ijkl}) = k_{\phi} (1 + \cos(n\phi_{ijkl} - \phi_s)) \quad (2.4)$$

where  $k_{\phi}$  is the force constant,  $n$  is the multiplicity and  $\phi_s$  is the phase shift. This functional form is often used in other force fields to describe proper dihedral angles. With a change in atom ordering, it is used in the OPLS-AA force field to express the out-of-plane bending, and can be employed, for example, to keep aromatic rings planar.

$V_{\text{LJ}}$  is the interaction energy between atoms  $i$  and  $j$  found at distance  $r_{ij}$ ,

$$V_{\text{LJ}}(r_{ij}) = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2.5)$$

with  $\epsilon_{ij}$  and  $\sigma_{ij}$  being the depth and the width of the 6-12 Lennard-Jones potential.

$V_{\text{Coulomb}}$  is the interaction energy between atoms  $i$  and  $j$ , found at distance  $r_{ij}$ , and bearing the partial charges  $q_i$  and  $q_j$ ,

$$V_{\text{Coulomb}}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2.6)$$

with  $\epsilon_0$  being the dielectric constant.

Each atom can be involved in one or more interactions of each type, and their effects on the atom add up. Summing over all atoms gives the potential energy of the system:

$$\begin{aligned} V_{\text{potential}} &= V_{\text{bonded}} + V_{\text{non-bonded}} \\ &= V_{\text{bonds}} + V_{\text{angles}} + V_{\text{dihedrals}} + V_{\text{impropers}} + V_{\text{LJ}} + V_{\text{Coulomb}} \end{aligned} \quad (2.7)$$

From the potential energy functions, forces can be obtained by derivation with respect to the coordinates. To prevent errors from numerical differentiation, forces are calculated by analytical derivation of the potential energy functions listed above (Equations 2.1-2.6). Summing up all forces acting on an atom gives the total force on that atom:

$$\mathbf{F}_i = \sum_j \mathbf{F}_{ji} \quad (2.8)$$

where  $\mathbf{F}_{ji}$  is the pairwise force acting on atom  $i$  due to the interaction with another atom  $j$ .

### 2.1.2 Integration

Once the forces acting on the atoms are known, Newton's equations of motions are integrated with respect to time, after choosing a small enough time increment:

$$\mathbf{F}_i = m_i \mathbf{a}_i = m_i \frac{d\mathbf{v}_i}{dt} = m_i \frac{d^2\mathbf{r}_i}{dt^2} \quad (2.9)$$

where  $\mathbf{F}_i$  is the total force acting on atom  $i$ ;  $m_i$ ,  $\mathbf{a}_i$ ,  $\mathbf{v}_i$  and  $\mathbf{r}_i$  are the mass, acceleration, velocity and position vector of atom  $i$ , respectively. The advance in time is often realized through a so-called leapfrog scheme, based on the Taylor expansions for velocity and position, for which the third-order and higher terms are considered negligible:

$$\mathbf{v}_i \left( t + \frac{\Delta t}{2} \right) = \mathbf{v}_i \left( t - \frac{\Delta t}{2} \right) + \mathbf{a}_i(t) \Delta t \quad (2.10)$$

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i \left( t + \frac{\Delta t}{2} \right) \Delta t \quad (2.11)$$

The integration provides a new set of atom positions and velocities which are used to start a new MD cycle. The positions at each integration step can be written out in a trajectory file and used later for analysis.

The choice of the time increment  $\Delta t$  (also known as time step) is very important. A large value allows MD simulations to reach time scales comparable to experimental ones (milliseconds or higher), but hides the high frequency fluctuations of light atoms. To accurately record changes in the position of hydrogen atoms or to represent non-equilibrium processes, the time increment should not exceed 1 fs ( $10^{-15}$  s). In practice, through the use of constraints like SHAKE<sup>26</sup> or LINCS<sup>27,28</sup> to restrict bonds between a heavy atom and hydrogen to their equilibrium distance, the time increment can be increased to 2-4 fs. In this work, we use a time increment of 1 fs for the MD simulations of graphene rupture (Chapter 3) and of 2 fs for the mechanical unfolding of proteins (Chapter 4).

### 2.1.3 Short range and long range interactions

In a molecular system containing only monoatomic gas molecules, each atom would be able to interact with all other atoms through non-bonded interactions. For a system composed of  $N$  atoms, the total number of interactions is  $N(N - 1)$ , which becomes quickly very large as  $N$  increases. In a typical biomolecular system, containing for example one or more proteins in water, some of the atoms form chemical bonds; as a consequence, the number of non-bonded interactions is slightly lower than  $N(N - 1)$ , but it retains a quadratic dependence on  $N$ . Both non-bonded potentials mentioned above (Lennard-Jones and Coulomb) are a function of the inverse of the distance between atoms (or a power of it, in case of the Lennard-Jones potential), so the interaction energy becomes lower as the distance increases. By introducing a *cut-off* distance beyond which the non-bonded interactions of each atom are neglected, the number of interactions is much reduced. For example, for a typical cut-off distance of 1 nm used with the OPLS-AA force field, the number of interactions of each atom is on the order of several hundreds. Compared with  $N - 1$ , which can be very large - hundreds of thousands or more - for multi-protein molecular systems, the decrease in the number of interactions can reach 2-3 orders of magnitude, leading to a very significant reduction in computational power and time required for calculations. Such interactions with atoms within the cut-off distance are called *short range* interactions.

While cut-off based schemes are very often used in MD simulations, the accumulation of neglected interactions can lead to artifacts. The Coulomb potential (which depends on  $r^{-1}$ ) decreases much slower with the increase in distance than the Lennard-Jones potential (which depends on  $r^{-6}$  and  $r^{-12}$ ). As a consequence, at a certain distance, the Coulomb potential could be small but non-zero, while the Lennard-Jones potential would be virtually zero. To recover the contribution to the total potential energy of these so-called *long range* electrostatic interactions, a grid (or mesh) based scheme has been introduced; initially based on an idea of Ewald<sup>29</sup>, it has been enhanced<sup>30,31</sup> and is now known under the

name of Particle Mesh Ewald (PME). This scheme replaces the individual atomic long-range interactions by interactions of point charges on a grid. Each point on the grid can have a charge determined by the surrounding atoms. These charges interact through a Coulomb potential, generating forces on the grid points which are then distributed to the individual atoms surrounding each grid point. In contrast to the interaction potentials mentioned before (Equations 2.1-2.6), the long-range electrostatic potential obtained through PME is calculated for the whole system at once and, thus, is not a pairwise potential.

The carbon atoms in our model of graphene do not carry any electrostatic charges, such that no Coulomb interactions are calculated (Chapter 3). In contrast, both short-range and PME-based long-range electrostatic interactions are calculated during the mechanical unfolding of proteins (Chapter 4), as molecular systems containing proteins and water are assigned partial charges on most of their atoms.

#### 2.1.4 *Periodic boundary conditions*

MD simulations are often performed on a single molecule of interest surrounded by solvent, for example a protein in water. To keep the simulation time reasonable, the total number of atoms is small compared to a similar experimental setup. As a consequence, effects of the boundary of the simulation system on the molecules of interest can lead to artefacts, such that physical properties derived from the simulation cannot be directly compared to the ones derived in experiment. To overcome this limitation, MD simulations often use *periodic boundary conditions (PBC)*, allowing the molecule of interest to experience the same forces it would in bulk solvent. The simulation is performed with the atoms placed in a box of regular shape with well defined boundaries; the unit box is virtually multiplied in all directions, such that the molecular system appears infinite. An atom moving beyond one boundary is replaced by an atom of the same type appearing near the opposite boundary, thus keeping constant the total number of atoms in the box. The most common box shape is a rectangular one (or cubic if all sides have equal lengths), but other shapes, like a truncated octahedron or a rhombic dodecahedron, might be more computationally efficient, reducing the number of solvent molecules needed to fill the box.

Although PBC make the molecular system appear infinite, the box cannot be made arbitrarily small - otherwise artifacts could appear due to a solute molecule interacting with its copy from a virtual neighbor box. Such a situation is prevented by placing the box boundaries at a distance larger than the non-bonded interaction cut-off from any atom of the solute in any direction. For a solute which is expected to change shape during the MD simulation, like in protein unfolding (Chapter 4), the box should be large enough to surround the solute in its most extended conformation.

2.1.5 *Temperature and pressure control*

The kinetic energy of a molecular system of  $N$  atoms can be written as:

$$E_{\text{kinetic}} = \left\langle \sum_{i=1}^N \frac{1}{2} m_i v_i^2 \right\rangle = \frac{3}{2} N k_B T(t) \quad (2.12)$$

where  $m_i$  and  $v_i$  are the mass and velocity of atom  $i$ ,  $k_B$  is the Boltzmann constant,  $T(t)$  is the temperature at time  $t$ , and the angle brackets surrounding the sum represent a time average. This equation links the instantaneous velocities of atoms to the temperature of the system and can be used to maintain a constant temperature by changing the atom velocities. In practice, the time averaging is performed over a period of several hundreds or thousands of time steps, such that the temperature can show small fluctuations around the desired value. Equation 2.12 also shows that the more atoms are present in the system the less significant the individual atom velocities are on the temperature fluctuations. A more complex and precise system for temperature control<sup>32,33</sup> adds an extra degree of freedom in form of a heat bath; energy is allowed to flow back and forth between the molecular system and the heat bath, such that the kinetic energy corresponds to the desired temperature as given by Equation 2.12.

Although a molecular system can be slowly brought to the desired temperature, a different procedure is very often employed: an initial set of random velocities which satisfies Equation 2.12 is assigned to the atoms, after which the system is allowed to relax while maintaining the temperature. Assigning different sets of initial velocities to the same molecular system allows it to move in different directions in the conformational space. For an even better sampling of the conformational space, not only the initial velocities but also the initial conformations should be different. We use such a scheme to generate different starting states to study the variation of rupture forces for graphene (Section 3.3) and of protein unfolding time under mechanical force (Section 4.2).

Similar to maintaining a constant temperature, it is also desirable to maintain a constant pressure in order to reproduce physiological or experimental conditions. In contrast to the temperature control, which only requires information about the atoms (their number, their masses and their velocities), the pressure control needs a well defined volume on which to act and, therefore, can only be used together with PBC. The pressure control relies on the virial theorem<sup>34,35</sup>, which connects the average kinetic and potential energies of a system of  $N$  atoms:

$$2\langle E_{\text{kinetic}} \rangle = -\langle E_{\text{potential}} \rangle \quad (2.13)$$

where the angle notation represents a time average. Based on this theorem, the following relation can be derived<sup>36</sup>:

$$pV = Nk_B T + \frac{\langle W \rangle}{3} \quad (2.14)$$

which links the pressure  $p$  and volume  $V$  of the system to the temperature  $T$  and the internal virial  $W$ ;  $k_B$  is the Boltzmann constant. The internal virial is written as:

$$W = \sum_{i=1}^N \mathbf{F}_i \cdot \mathbf{r}_i \quad (2.15)$$

where  $\mathbf{F}_i$  is the force acting on atom  $i$ , the location of which is given by the position vector  $\mathbf{r}_i$ . After calculating the right side of Equation 2.14 and given the desired pressure, the volume of the PBC box can be determined. As the temperature and internal virial vary from one time step to the next, the dimensions of the PBC box fluctuate continuously during the MD simulation.

MD simulations can be classified based on the presence or absence of temperature and pressure control. In addition to temperature and pressure, other measures related to the system as a whole can be kept constant during the simulation. The great majority of MD simulations are performed with a constant number of atoms  $N$ . In some cases the total energy  $E$  of the molecular system is conserved, such that any change in potential energy is translated into a change in kinetic energy and *vice versa*. Typically, several such measures are kept constant simultaneously, forming a so-called *ensemble*. In an *NVE* ensemble, the number of atoms, the volume and the energy do not change over the course of the simulation. In an *NVT* ensemble, the temperature is maintained constant through a coupling to a heat bath, and energy is exchanged between the molecular system and the heat bath. In an *NpT* ensemble, both the pressure and temperature are controlled; the volume of the system changes to maintain the constant pressure. The *NVE* and *NVT* ensembles are normally used to allow the system to relax and reach an equilibrium state, during the so-called *equilibration phase*; the molecular system is considered to have reached an equilibrium state when one or more observable measures remain relatively constant in time, with only small fluctuations. Such an equilibrium state is then used as the start for simulations in an *NpT* ensemble, during which data is collected and analyzed - the so-called *production phase*. To reproduce experimental conditions of temperature and pressure, MD simulations in both the *NVT* and *NpT* ensembles are used in this work.

### 2.1.6 Applying an external force

As mentioned in Chapter 1, biological processes are often associated with or investigated by the application of mechanical forces on the molecules of interest. When using MD simulations to study, for example, the process of unbinding of a small ligand from a protein receptor, the probability of observing the binding process is very low. By applying a small mechanical force between the two molecules, directed such that the ligand is pulled away from the binding pocket on the surface of the protein, the unbinding probability increases significantly<sup>37,38</sup>. The magnitude of the applied force has to be small enough to prevent introducing artifacts in the structure or the sequence of events making up the process under study. For example, a too high force could induce structural distortions, leading to an unnatural contact surface between the protein and ligand. MD simulations in which mechanical forces are applied on the molecules of study are called Force-Probe Molecular Dynamics<sup>37</sup> or Steered Molecular Dynamics<sup>38</sup>.

In order to closely reproduce experimental conditions, the magnitude, direction and/or application point of an external force can be kept constant in MD simulations. The force can be applied on an individual atom or on a group of atoms, which could also be a whole molecule. A force applied on a single atom is directly added to the atomic force (Eq. 2.8); a force applied on a group of atoms acts on the center of mass (COM) of the group and is distributed in a mass-weighted fashion to the individual atoms. In some studies, the force application point has to reach a certain position which can be constant or move with constant velocity or acceleration. To accomplish this, a harmonic potential is used to reduce the distance between the current and the desired position, the concept being similar to the bond potential (Equation 2.1).

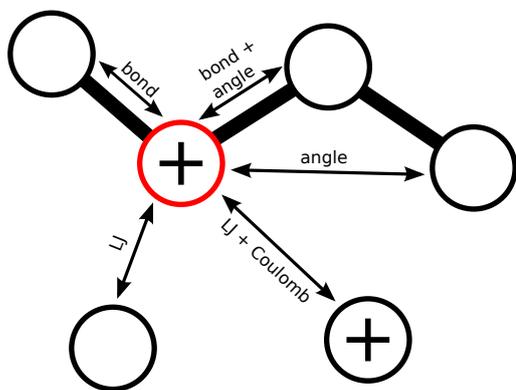
An example of external force application is the calculation of the free energy of ligand binding. This is often performed through the so-called *umbrella sampling*, in which the ligand and receptor are kept at various distances from each other by an *umbrella potential* - a harmonic potential which penalizes any deviation from the desired distance. The force between the ligand and the receptor needs to change magnitude and direction to counteract the random movement of the molecules, and is analyzed to derive free energy changes of the process. Another example is the simulation of AFM experiments (Section 1.1), which can be performed with either a constant force or a constant velocity of the indenter tip. The indenter tip has a well defined motion, so the direction of the force remains unchanged during the experiment.

A mechanical force was applied in most of the MD simulations carried out as part of this work. For the study of graphene rupture (Chapter 3), the force was applied on a group of atoms - a sphere resembling an indenter tip - which moves in one direction with con-

stant velocity. To investigate protein unfolding under mechanical force (Chapter 4), a force with constant magnitude and direction was applied on individual atoms - the  $C_{\alpha}$ -atoms of terminal protein residues.

## 2.2 TIME-RESOLVED FORCE DISTRIBUTION ANALYSIS

As mentioned in Section 2.1.1, atoms interact during MD simulations through forces derived from the potential energy functions. During an integration step, the movement of an atom  $i$  is determined by the total force acting on it,  $F_i$ , obtained as a sum of pairwise forces between atom  $i$  and any atom  $j$  with which it interacts (Equation 2.8 and Figure 2.1). Although calculated during each simulation step, the individual pairwise forces are lost due to the summation.



**Figure 2.1:** Various types of interactions of an atom (red) with other atoms (black) are added up to determine the total force. Thick black lines represent bonds between atoms. Interactions between atoms not involving the central one (red) are not shown.

The typical outcome of an MD simulation is a collection of three-dimensional coordinates, indicating the positions of the atoms, but missing any kind of relation between them. In contrast, pairwise forces between atoms or residues describe their interactions independent of their actual positions and, thus, represent internal coordinates of the molecular system. Pairwise forces are also a more sensitive measure for changes that happen during an MD simulation. For example, two pairwise forces of equal magnitude but opposite directions acting on a atom will result in the atom maintaining its position, as the total force

exerted on it is zero. In this case, analysis of the pairwise forces allows recovering the individual interactions of the atom, while the widely used coordinates-based methods will miss them entirely. Based on this principle, it was possible to highlight low-amplitude yet functionally important motions, such as those in a stiff protein core<sup>39</sup>, to explain the mechanical robustness of immunoglobulin domains<sup>40</sup> and to reveal the pre-stress in proteins<sup>41</sup>. The method used in these studies, called Force Distribution Analysis (FDA)<sup>42</sup>, allows examining the internal forces and their spatial distribution from MD simulations. FDA has been applied so far only to averaged data from equilibrium simulations, leading to a static view of the force distribution. However, many biomolecular systems or other complex macromolecules can visit a broad range of conformational states in equilibrium conditions. Fur-

thermore, external perturbations, such as interactions with a ligand or the application of a mechanical force, can change the rates of transition between these states or open the possibility of visiting different states. Each state can be characterized not only by the three-dimensional coordinates of the atoms, but also by the spatial distribution of internal forces. In order to study transitions between conformational states, it is therefore necessary to follow the evolution in time of the internal forces. This is difficult to achieve with FDA and prompted a new implementation.

We introduce here the **Time-Resolved Force Distribution Analysis (TRFDA)**, which extends FDA by adding a temporal component to enable the analysis of pairwise forces during conformational changes. Similar to FDA, atomic pairwise forces are computed for all types of bonded interactions as well as for short-range Coulomb and Lennard-Jones potentials (Section 2.1.1). Long-range electrostatic interactions computed on a grid (Section 2.1.3) cannot be decomposed in atomic pairwise forces and are thus not included in the analysis. Also similar to FDA, a trajectory is first obtained from an MD simulation, and is then analyzed by using the TRFDA code. The implementation was written from scratch for GROMACS<sup>43</sup> 4.5.3 and contains over 3500 lines of code. For visualization of the pairwise forces and punctual stress (Section 2.3.3), VMD<sup>44</sup> plugins are provided. The entire software distribution is available as open source from the project website:

<http://code.google.com/p/force-distribution-analysis/>

Apart from focusing on the evolution of pairwise forces, TRFDA also includes several notable improvements over FDA, like a force decomposition for 3- and 4-body potentials, internally computed pairwise forces between residues, the possibility to sum up all forces between the same pair of atoms, two ways of converting vector to scalar forces, the ability to define two groups of atoms between which pairwise forces are calculated, a memory-efficient internal data organization, and data output in easy-to-parse file formats. The multi-body force decomposition, residue pairwise forces, and vector to scalar transformations are described below; the others - of more technical nature - are detailed in our publication describing TRFDA<sup>45</sup> and in the TRFDA manual, available from the project website.

### 2.2.1 *Decomposition of forces from 3- and 4-body potentials*

For two-body potentials such as bonds, Coulomb and Lennard-Jones potentials, the interaction force derived from the potential energy function represents the pairwise force between the two atoms involved. For many-body potentials, which describe the interactions between more than two atoms, the atomic forces need to be decomposed into pairwise forces. TRFDA introduces a complete decomposition of the forces resulting from 3- and 4-body potentials (angle, dihedral angle, cross bond-bond, cross bond-angle), described in

the Appendix A. The force decomposition is performed on the atomic forces calculated from the various interaction potentials and is therefore independent of the form of these potentials. For example, the same decomposition can be applied to the two different proper dihedral angle potentials in GROMACS (periodic and Ryckaert-Belleman, Section 2.1.1). The resulting pairwise forces can have any direction, *i.e.* they do not align with the distance vector between the two atoms involved. To accommodate this, all pairwise forces are stored and handled internally as vectors.

This decomposition rectifies the shortcomings of the previous FDA implementation, which used approximations for computing pairwise forces resulting from 3- or 4-body potentials. For an angle formed by atoms  $i$ ,  $j$  and  $k$  ( $j$  placed between  $i$  and  $k$ ), it considered that no pairwise forces act on atom  $j$ , and assumed that the pairwise force between atoms  $i$  and  $k$  acts along the distance vector between the two atoms. A similar decomposition has been used for a dihedral angle formed by four atoms. In contrast to these approximations, the decomposition used in TRFDA correctly reproduces the distribution of angle and dihedral forces in a molecule.

### 2.2.2 Internally computed pairwise forces between residues

Computing pairwise forces between residues allows a significant decrease in storage and computational cost for analysis, while providing a mapping of the interactions on the primary and possibly also secondary structure of a protein. The examination of pairwise forces between residues can also be used as a tool in the development of residue-level coarse grained models<sup>46</sup>. The pairwise force representing the interaction between residues  $r_i$  and  $r_j$  acts on the centers of mass of the two residues and is calculated as:

$$\mathbf{F}_{r_i,r_j} = \sum_{i \in r_i, j \in r_j} \mathbf{F}_{ij} \quad (2.16)$$

where  $i$  is an atom of residue  $r_i$ , and  $j$  is an atom of residue  $r_j$ , with  $r_i$  different from  $r_j$ . TRFDA computes internally pairwise forces between residues; when only forces between residues are of interest, the amount of memory required is significantly reduced, and writing out pairwise forces between atoms is entirely avoided. This constitutes a significant improvement over the previous implementation of FDA, which stored in memory a large number of pairwise forces between atoms, saved them in a large file, and subsequently read them by a standalone tool which calculated residue pairwise forces.

TRFDA treats equally the atomic and residue pairwise forces with respect to further processing and output. For example, the same vector to scalar transformations can be applied to both pairwise forces between atoms and pairwise forces between residues, punctual

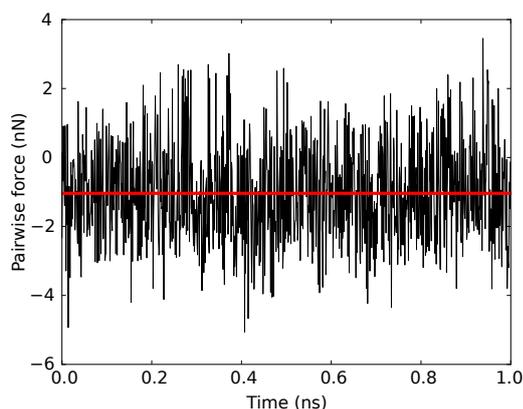
stress (Section 2.3.3) can be calculated both per atom and per residue, and the output file formats are the same for data referring to atoms or to residues. Furthermore, the same graphical representations can be applied to both atom and residue pairwise forces or punctual stress.

### 2.2.3 Vector to scalar transformations

Although pairwise forces are calculated and stored internally only in vector form, TRFDA can write them out either as vector or as scalar values. A scalar pairwise force is computed as the magnitude of the vector pairwise force, similar to FDA, or as the magnitude of the pairwise force projected onto the distance vector between the two atoms<sup>41</sup>. In either case, the scalar value carries a sign indicating whether it is a repulsive (plus) or attractive (minus) force. If the angle between the vector pairwise force and the distance vector between the atoms is in the range  $(-\pi/2, \pi/2)$ , the pairwise force is considered attractive. If the angle between the vector pairwise force and the distance vector between the atoms is in the range  $(\pi/2, 3\pi/2)$ , the pairwise force is considered repulsive. If the vector pairwise force is perpendicular to the distance vector, the pairwise force can be considered neither attractive nor repulsive and is set to zero.

### 2.2.4 Noise reduction

In most MD simulations, the atom velocities continuously vary, leading to fluctuations in atom positions and, consequently, in pairwise forces. Very often, these fluctuations appear as changes between positive and negative values, corresponding to changes between repulsive and attractive forces, resulting in a high noise of the force signal (Figure 2.2). Even more, the magnitude of these fluctuations can be as large, or even larger, than other forces acting in the system; for example, it is difficult to distinguish an external force (Section 2.1.6). To reduce this noise, pairwise forces can be averaged in time, in space or between several simulations. A time average can be obtained through a block based approach (implemented



**Figure 2.2:** Variation of the pairwise force in a carbon-carbon bond in graphene during an equilibrium simulation. The average value of  $-1.04 (\pm 1.41)$  nN is represented in red. This non-zero value is the expression of the background stress (Chapter 3).

in TRFDA) or a moving window approach. For an averaging in space, the molecular system should maintain a relatively constant geometry and be relatively homogeneous, with equivalent interactions found at different places in the system; this is the case for graphene (Chapter 3) or a lipid bilayer. Finally, averaging can be performed over simulations which sample the same conformational space (Section 2.1.5), the result describing the pairwise forces for the entire ensemble. Apart from averaging, noise can also be reduced by applying classical signal processing techniques, for example based on the Fast Fourier Transform or convolution.

In this work, we use all the averaging methods mentioned above to reduce the fluctuations in pairwise forces. As part of the circular stress calculation in graphene (Section 2.3.4), pairwise forces are averaged in space, namely over atom pairs found at the same distance from the center of the sheet. Averaging between several simulations sampling the same conformational space and averaging over time are performed for the study of pairwise forces during the mechanical unfolding of proteins (Chapter 4).

### 2.3 STRESS AT ATOMIC LEVEL IN MD SIMULATIONS

Mechanical stress is defined as the ratio between the force and the area on which the force acts, and can be applied at both macro- and microscopic scales. The study of mechanical stress has seen significant advances over time. Nowadays, the distribution of stress can be very easily and accurately predicted with the help of computer simulations, using Finite Elements (FE) models. Such models can be built for objects of very different sizes, from screws of few millimeters, to cars of few meters, to airplanes of tens of meters, to bridges of hundreds of meters or kilometers. Calculations of stress can be performed both in equilibrium conditions and under some external influence which manifests as a mechanical deformation. Such calculations are very important in ensuring that structural elements do not fail when they are exposed to an external perturbation; for example, a bridge should not collapse under the load of passing cars, an airplane wing should not break due to wind, etc. Stress calculations are not only important for choosing the right geometry of structural elements, but also for choosing the right material for them as each material has a maximum stress it can sustain before failing. The stress might not be equally distributed in a structure and the material starts failing only in places where the stress becomes too high. However, the failure might not be immediately visible and in many cases it starts with a crack at atomic scale. Therefore, it is highly interesting to have methods for the analysis of the stress distribution at atomic level and for the prediction of crack formation and propagation. Such methods are not limited to industrial materials, but should be applicable to molecules of biological interest as well.

Experimentally measuring how forces or stresses are distributed at atomic level is currently impossible, but studies can be carried out on FE and molecular models. TRFDA, developed as part of this work (Section 2.2), can easily be used to examine the internal forces and their spatial distribution. Computing the atomic level stress is however not straightforward, as the area on which atomic forces act is not well defined. The following sections describe two previous representations of stress, the atomic virial stress<sup>47</sup> and the local pressure<sup>48</sup>, and introduce two new representations, the atomic punctual stress and the circular stress.

### 2.3.1 Atomic virial stress

To represent stress at atomic level in MD simulations, the virial theorem<sup>34,35</sup> was used to define a per atom virial stress<sup>47</sup>. The pressure in a system of volume  $V$  is determined by the internal virial  $W$  defined above (Section 2.1.5) which takes into account the atomic forces  $F_i$ . Similarly, the virial stress for an atom occupying the volume  $V_i$  can be determined from the pairwise forces  $F_{ji}$  between the atom  $i$  and any other atom  $j$ .

Due to the calculation as  $F_i \cdot r_i$  (Equation 2.15), the virial stress  $\sigma$  is stored as a tensor with 6 components: 3 components (xx, yy and zz) represent the normal stress, while the rest (xy, yz and zx) represent the shear stress. The von Mises yield criterion can be used to reduce the tensor form to a single value:

$$\sigma = \sqrt{\frac{(\sigma_{xx} - \sigma_{yy})^2 + (\sigma_{yy} - \sigma_{zz})^2 + (\sigma_{zz} - \sigma_{xx})^2 + 6(\sigma_{xy}^2 + \sigma_{yz}^2 + \sigma_{zx}^2)}{2}} \quad (2.17)$$

Instead of the typical definition of stress as force over area, this representation uses an equivalent one of force times distance, or energy, over volume. The volume occupied by an atom is not well defined and therefore omitted from calculations, such that the atomic virial stress is expressed in terms of energy; a true stress can only be obtained once the atomic volumes are known. As the equations above describe any molecular system, the atomic virial stress can be regarded as a general purpose atomic stress representation.

### 2.3.2 Local pressure

The local pressure concept<sup>48</sup> was introduced to overcome difficulties in defining geometric properties like area and volume at atomic level. It uses a three-dimensional grid overlapped onto the molecular system and calculates stress from the interatomic forces acting on the faces of the grid cells. The grid cells need to be small enough such that any two atoms in

the molecular system are found in two different cells; interactions between atoms found in the same cell do not “cross” any grid cell face and thus are not taken into consideration. As both the forces and the area on which they act are well defined, the local pressure uses units of force over area, like a true stress.

This approach works well when the molecular system has a regular shape and is mostly aligned to the grid, as is the case for a lipid bilayer. It is however very hard to apply for a protein of irregular shape, moving and possibly deforming during an MD simulation. A further disadvantage is that the local pressure offers information about the stress at an absolute location, given by the grid, but not on how different atoms, residues or groups interact with each other. For this reason, the variations of stress during conformational changes are very hard to track. This limits its usability to the analysis of molecules or ensembles of molecules with a regular shape, which does not change during the MD simulation.

### 2.3.3 Punctual stress

The mechanical stress measures the resistance of the material against physical deformation or rupture upon applying an external force. When several external forces act simultaneously on the material, their effects accumulate; the total stress becomes a sum of all the individual stresses caused by these forces. Similarly, we can define the stress on an atom as the sum of stresses induced by all pairwise forces acting on the atom. However, the area of action of atomic pairwise forces is unknown and therefore ignored. The stress on an atom  $i$  becomes a sum of the absolute values of scalar pairwise forces acting on it:

$$\sigma_i = \sum_j |F_{ji}| \quad (2.18)$$

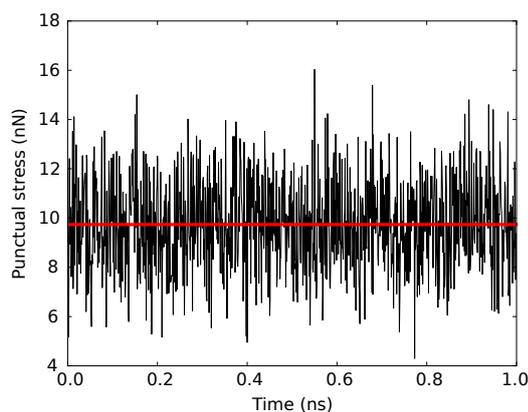
We denote the stress defined in Equation 2.18 as *punctual stress*, emphasizing the action of the force on a dimensionless point instead of an area, which is here ill-defined. As a sum of forces, the punctual stress is expressed in units of force.

In equilibrium conditions, the pairwise forces should converge to non-zero values, and so should the punctual stress, as a sum of convergent quantities. In contrast, the sum of vector pairwise forces acting on a single atom, typically computed in MD simulations, averages to zero in the same conditions. This important property of the punctual stress allows segregating the presence of force from the effect it usually produces - movement of atoms. To illustrate this, we can consider two examples: one in which two pairwise forces of equal magnitudes and opposite directions act on an atom and the second one in which the pairwise forces are zero. In the first example, the sum of the vector forces is zero, while the punctual stress is equal to the sum of the magnitudes of the forces, and thus is non-zero.

In the second example, both the sum of vector forces and the punctual stress are zero. The atom does not move in either example and this would imply that the two examples are equivalent. The punctual stress can however differentiate between them, describing the presence of forces even in the absence of movement. This is a further argument for calling this measure “stress” despite its expression as a sum of forces: just as the mechanical stress is able, for example, to represent the presence of internal forces in an airplane wing under its own weight, so is the punctual stress able to represent the presence of pairwise forces on a non-moving atom, in the first example above.

In equilibrium conditions, an atom typically fluctuates around an average position, and the individual pairwise forces acting on it often change value and sign. As a sum of pairwise forces, the punctual stress on the atom can also vary (Figure 2.3).

As opposed to the analysis of pairwise forces, which gives a very detailed view of their distribution in or between molecules, the punctual stress expresses in a simple way where pairwise forces accumulate, and allows the detection of atomic-level “hot-spots”. Once these locations are identified, a more detailed analysis of pairwise forces may be performed.



**Figure 2.3:** Variation of the punctual stress on a carbon atom of graphene, during an equilibrium simulation in vacuum. The non-zero average value (red) represents the background stress (Chapter 3).

#### 2.3.4 Circular stress

In both biological and non-biological environments, we can find molecular systems for which the morphological and physical properties are very similar for two of the three geometrical dimensions and significantly different for the third dimension. A lipid bilayer represents such an example, where the system is heterogeneous in the dimension perpendicular to the bilayer (direction of the thickness of the bilayer), but mostly homogeneous in the other two dimensions. The so-called two-dimensional materials, like graphene (Chapter 3), hexagonal boron nitride or sheets of transition metal oxides, represent a special case where the material is one atom thick in the dimension perpendicular to the sheet.

An AFM cantilever, pushing on such a molecular system, generates a localized force which dissipates radially from the application point, a property which can be exploited to estimate the stress distribution (Chapter 3). We can define concentric cylinders with the

center on the force application point and increasing radii, and with height equal to the thickness of the molecular system. Based on their position with respect to each cylinder, the atoms are split between the inside and the outside of it. For a pair of interacting atoms found on the opposite sides of the cylinder, the pairwise force has a component normal to the lateral surface of the cylinder. The stress  $\sigma$  can then be defined as the sum of all normal components of pairwise forces between atoms on the inside and atoms on the outside of the cylinder divided by the lateral area of the cylinder:

$$\sigma = \frac{\sum_{i,j} F_{ij}^{\perp}}{2\pi r h} \quad (2.19)$$

where  $F_{ij}^{\perp}$  is the component of the pairwise force between atoms  $i$  and  $j$  normal to the surface of the cylinder, with atoms  $i$  and  $j$  on opposite sides of the cylinder;  $r$  and  $h$  are the cylinder radius and height, respectively. This definition uses the typical expression of stress of force over area.

For two-dimensional materials like graphene, the thickness of the material is not well defined. Equation 2.19 can be modified to ignore the height of the cylinder:

$$\sigma^{2D} = \frac{\sum_{i,j} F_{ij}^{\perp}}{2\pi r} \quad (2.20)$$

In consequence, the stress is in this case expressed as force over length.

### 2.3.5 Comparison of atomic stress definitions

As can be seen above, the definition of atomic level stress is not straightforward, the main difficulty being the definition of an area of action of force at atomic level. To sidestep this difficulty, the local pressure and circular stress use geometrical constructs fixed in space (a grid or a cylinder, respectively) for which an area can be defined easily, and calculate forces acting across these areas. The fixed location in space of these constructs represents a significant disadvantage, limiting their usability to molecular systems which are themselves fixed in space (*e.g.* a circular graphene sheet clamped around the edge) or are homogeneous along the coordinate of interest (*e.g.* a lipid bilayer). Assigning such stress to an atom or to a basic unit of the molecular system (*e.g.* a lipid head) is performed indirectly and is imprecise, requiring first a localization of the atom or basic unit relative to the geometrical construct used. The main advantage is that these definitions of stress include both a force and an area of action of the force, being therefore expressed as a true stress.

In contrast, the atomic virial stress and punctual stress are directly assigned to atoms; the punctual stress can additionally be assigned to basic structural units, like aminoacids. These stresses can be calculated independent from the movement of atoms during an MD simulation and, therefore, can be applied to any molecular system. Both these stresses deviate from the expression of a true stress as force over area, due to the difficulty of defining geometrical properties at atomic level: the punctual stress uses units of force as the area of action of pairwise forces is ill-defined, while the atomic virial stress uses units of energy as the volume of an atom is also ill-defined. Even though these expressions of stress use non-standard units, they are capable of expressing the relative accumulations of forces in the molecular system.



# 3

---

## GRAPHENE RUPTURE

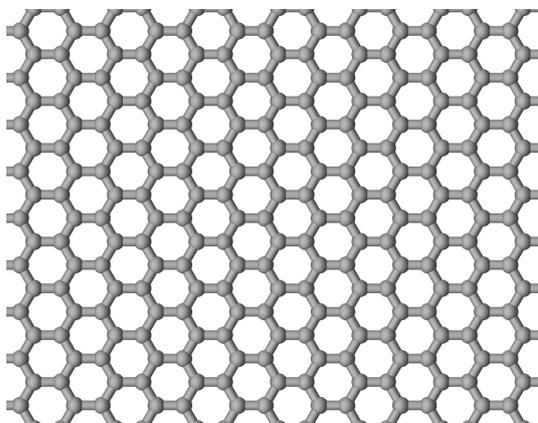
---

### 3.1 INTRODUCTION

Graphene is a single layer of graphite, which is composed of carbon atoms tightly packed into a two-dimensional hexagonal lattice (Figure 3.1). Although its existence was proposed several decades ago<sup>49</sup>, a method for isolating graphene has been found only recently, allowing its properties to be studied for the first time<sup>50</sup>. Since then, the interest in graphene and carbon nanotubes (CNTs) - graphene sheets rolled as cylinders - has grown exponentially. Due to its many favorable physical and chemical properties, graphene has been called “the wonder material” and is considered for applications in many fields<sup>51</sup>.

The interest in graphene has led to a large body of theoretical and experimental results. Theoretical studies cover scales ranging from quantum mechanical (QM) calculations, through classical MD simulations, to Finite Elements (FE) models. QM calculations are computationally expensive, but take into account the electronic degrees of freedom, allowing investigations of electron delocalization and excitation<sup>52,53</sup>, and chemical reactivity. FE models can bridge the nano- and macroscopic scales, but miss the atomic details,

such that a complete picture about the material rupture and crack propagation cannot be obtained.



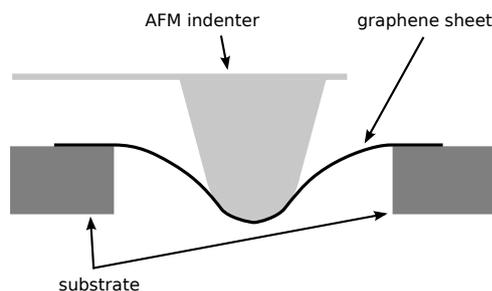
**Figure 3.1:** Carbon atoms in a single layer graphene sheet.

Before experimental studies were possible, various models have tried to predict properties of graphene and CNTs. For MD simulations, the pairwise potentials typically used for proteins (Section 2.1.1) have been considered insufficient and, therefore, specific multi-body potentials have been developed<sup>54-60</sup>. To correctly represent changes in electron delocalization due to the displacements of atoms in carbon-based compounds, these potentials rely on the concept of bond order. They take into account not only the bond but also its environment, including in some cases long-range interactions<sup>58</sup>. Although successfully used to study physical properties of graphene and CNTs, the bond order potentials have several disadvantages with regard to efficiency and complexity of implementation, as discussed in further detail in our publication<sup>61</sup>.

Graphene and CNTs might also find biological applications<sup>51</sup>. It is therefore not surprising that mixtures of them with biomolecules have already been investigated by means of MD simulations<sup>62-64</sup>. Such studies have modeled the graphene and CNTs either as rigid bodies or through harmonic pairwise potentials, as the software packages used in biomolecular simulations<sup>43,65,66</sup> do not typically include the carbon-specific multi-body potentials. These approximations are valid because biomolecules are much softer than graphene and, thus, mechanical interactions do not stretch the graphene C-C bonds far way from the equilibrium bond length. However, a harmonic bond potential is not appropriate to describe the elongation or breaking of the C-C bond which could result from mechanical stress, like high pressure or an external force. A more realistic model for such situations uses the Morse potential, which is able to reproduce the energy convergence towards the dissociation energy upon bond elongation. It was already used to represent C-C interactions in graphene and CNTs<sup>67</sup>, and is only slightly more computationally expensive than a harmonic potential. In contrast to the carbon-specific multi-body potentials, the Morse potential is already implemented or can be straightforwardly added to biomolecular software packages. Building on these advantages, the initial part of our work focuses on the development of a Morse-based potential, able to reproduce important mechanical properties of graphene, but at the same time being simple to implement and offering a high computational efficiency.

Before experimental studies were possible, various models have tried to predict properties of graphene and CNTs. For MD simulations, the pairwise potentials typically used for proteins (Section 2.1.1) have been considered insufficient and, therefore, specific multi-body potentials have been developed<sup>54-60</sup>. To correctly represent changes in electron delocalization due to the displacements of atoms in carbon-based compounds, these potentials rely on the concept of bond order. They take into account not only the bond but also its environment, including in some cases long-range interactions<sup>58</sup>. Although successfully used to study physical properties of graphene and CNTs, the bond order potentials have several disadvantages with regard to efficiency and complexity of implementation, as discussed in further detail in our publication<sup>61</sup>.

High mechanical resistance is one of the many intriguing properties of graphene. By using an Atomic Force Microscope (AFM) indenter with a diamond tip (Figure 3.2), Lee *et al.*<sup>1</sup> obtained load-indentation curves and determined the rupture force for defect-free graphene sheets, finding graphene to be one of the strongest materials known



**Figure 3.2:** Schematic representation of the AFM indentation of a graphene sheet.

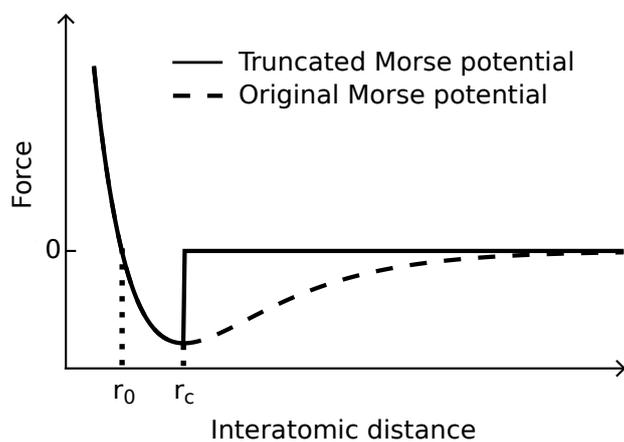
to date. The resistance of the single layer of carbon atoms to very large out-of-plane deformation forces is remarkable and has already led to several studies. Previous models, using FE<sup>68</sup> or MD<sup>69</sup> methods, have been derived from AFM experiments<sup>1,70</sup> by assuming a point load on the graphene sheet and, therefore, not taking into account the influence of the indenter on the circular plate. Furthermore, the studies focused only on the deformation of the sheet, neglecting the rupture aspects. Other models have investigated the mechanical rupture of graphene due to in-plane deformations<sup>71,72</sup>. In consequence, it is largely unknown how graphene can withstand the enormous forces probed experimentally under out-of-plane deformation, and how the non-linear elastic deformation of graphene under the indenter is related to the breaking of individual bonds.

The present work attempts to answer these questions by providing an atomic-level view of stress variation during graphene deformation until rupture. In MD simulations, we reproduce closely the experimental conditions, allowing us a direct comparison and validation of the results. We first introduce a simple modification of the Morse potential which allows monitoring bond breaking in graphene, and we parametrize it by fitting to high level quantum mechanical calculations on graphene patches. We then use this potential to describe the interactions in a circular graphene sheet and model the indenter as a sphere built from discrete atoms. Further, we validate our model by comparing the rupture forces for a range of indenter velocities and radii as well as graphene sheet radii to the experimental values. From force-indentation profiles obtained from MD simulations, we derive a two-dimensional elastic modulus, which we use for further validation. Finally, we follow the stress distribution throughout the graphene sheet during the indentation process until rupture by means of TRFDA and virial atomic stress calculations, and characterize the probability and location of material rupture.

## 3.2 A TRUNCATED MORSE POTENTIAL

The harmonic potential (Equation 2.1) is typically used to model bonds between atoms in proteins and other biomolecules. Elongating such a bond results in a linear increase of the attractive force, which continues even when the interatomic distance becomes large (nanometers or more). Such large bond lengths are however unrealistic, and bond breaking should occur before reaching them. In contrast, the Morse potential represents the attractive forces decreasing exponentially towards zero as the interatomic distance increases, offering a more realistic model of bond breaking. The Morse potential has already been used to study the fracture of graphene and CNTs<sup>67</sup>, where it has been shown that, for small initial defects, the inflection point of the potential and much less the depth of the potential (*i.e.* the dissociation energy) determines the maximum mechanical resistance of the material.

The inflection point of the Morse potential corresponds to the maximum attractive force. As this is the point where the bond resists the most to breaking, it can be defined as the *critical point*, and the corresponding bond length as the *critical bond length*,  $r_c$ . This definition implies that the bond can be considered broken for interatomic distances larger than  $r_c$ . Although the attractive force derived from the Morse potential is very close to zero at large bond elongations, it can still act as a restoring force, especially at bond lengths only slightly larger than  $r_c$ , leading to unrealistic bond reformation.



**Figure 3.3:** The variation of force with respect to the interatomic distance for the original and truncated Morse potentials.  $r_0$  represents the equilibrium bond length,  $r_c$  represents the critical bond length.

To prevent broken bonds from reforming, we introduce a slight modification to the Morse potential. For bond lengths lower than or equal to  $r_c$ , it retains the form of the original Morse potential, while for bond lengths higher than  $r_c$ , the potential remains constant, such that the force becomes zero (Figure 3.3). As a consequence, a bond elongated beyond its critical length is unable to converge on its own towards the equilibrium bond length, preventing unrealistic reformation of

bonds stretched to several nanometers. The attractive force can become non-zero again only if thermal fluctuations or other interaction potentials act on the two atoms previously

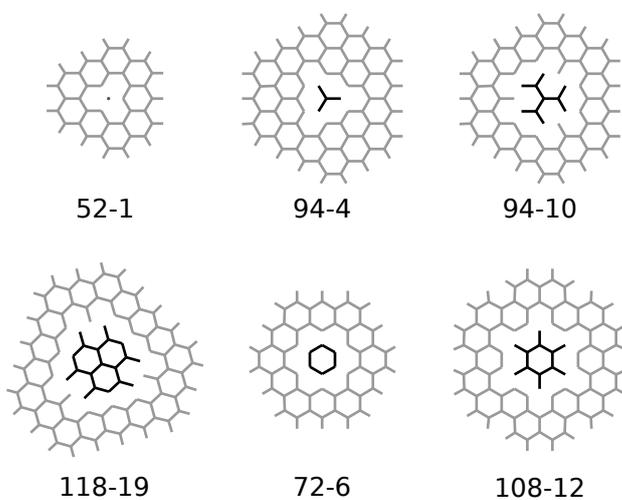
forming the bond and bring them closer than  $r_c$ , equivalent to a bond reformation. To prevent this from happening, our implementation of the truncated Morse potential reports the bond as broken when it crosses the critical bond length, and can optionally stop the MD simulation. We note that the Morse potential, in its original form as well as in the truncated form introduced here, is unable to model bond reformation between different C atoms or between C atoms and other atomic species, without significant code modifications in the MD software packages. Carbon-specific potentials (AIREBO<sup>57</sup>) or general reactive force fields (ReaxFF<sup>59</sup>) are needed in these cases, with the associated increase in computational time and complexity of implementation.

The Morse potential is already present in GROMACS<sup>43</sup>. For two atoms  $i$  and  $j$ , it has the form:

$$V_{\text{Morse}}(r_{ij}) = D_{ij}[1 - e^{-\beta_{ij}(r_{ij}-b_{ij})}]^2 \quad (3.1)$$

where  $D_{ij}$  is the dissociation energy,  $\beta_{ij}$  is a constant defining the steepness of the energy well and  $b_{ij}$  is the equilibrium distance.  $\beta_{ij}$  can be derived from the force constant for the harmonic bond potential and  $D_{ij}$ <sup>73</sup>.

$D_{ij}$  is typically obtained from experimental data or QM energy profiles. It corresponds to the variation of one bond potential with the interatomic distance while keeping constant all other interactions. It is however impossible to carry such measurements in graphene, given its network of identical bonds; for example, the movement of a single atom disturbs 3 bonds and 3 aromatic rings simultaneously. We therefore use a two-step approach for calculating  $D_{ij}$ : gradually shifting a set of C atoms out of the graphene plane (Figure 3.4) in QM calculations, and using the obtained energy profiles as reference for force field calculations, in which  $D_{ij}$  acts as the fitting parameter. Thus, instead of trying to identify the



**Figure 3.4:** Geometry of the graphene patches used for gradually shifting carbon atoms (black) out of the graphene plane (gray), identified by two numbers: the total number of atoms (including hydrogen atoms) and the number of carbon atoms shifted from the graphene plane. The sets of shifted atoms are defined by adding layers of carbon-carbon bonds to a core consisting of a carbon atom (52-1, 94-4, 94-10 and 118-19) or an aromatic ring (72-6 and 108-12).

energy difference corresponding to the elongation of an individual aromatic C-C bond, we consider the energy change of the whole atomic system as a reference to which we fit the Molecular Mechanics (MM) energy profiles of the same system.

We carried out the QM calculations using Gaussian09 with the 6-31+G\* basis set and several methods covering Density Functional Theory, hybrid and self-consistent charge density functional tight binding. As the bond is considered broken beyond the critical length and the potential is truncated, we fit MM potential energies to the QM energy profiles only for the range of bond lengths between the equilibrium bond length and  $r_c$ . Consequently, the truncated Morse potential reproduces only the increase in energy from zero to the inflection point. The fitting procedure uses the least square method and accounts for the fact that  $r_c$  depends on  $D_{ij}$ , so a different number of points from the QM energy profiles may be taken into account for each  $D_{ij}$ .

One of the most commonly used force fields in biomolecular MD simulations is OPLS-AA<sup>25</sup>. To obtain a truncated Morse potential for graphene compatible with it, we based our initial parameters on those describing groups with aromatic rings in OPLS-AA. We used the force constant and the equilibrium length for an aromatic C-C bond,  $392459.2 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  and  $0.140 \text{ nm}$ , respectively. We also used the corresponding OPLS-AA parameters for the description of angles, proper and improper dihedral angles, and for the Lennard-Jones potential. From the fitting procedure, we obtained a  $D_{ij}$  value of  $805 \text{ kJ mol}^{-1}$ . The corresponding critical bond length is  $0.184 \text{ nm}$ , and the maximum force to which the bond can resist is  $6285.6 \text{ kJ mol}^{-1} \text{ nm}^{-1}$  (or  $10.43 \text{ nN}$ ). We note that  $D_{ij}$  is only used as a bond-characteristic parameter in the fitting procedure, and therefore primarily reflects the slope in the potential up to the inflection point. In consequence,  $D_{ij}$  should not be directly compared to the dissociation energy known for an aromatic bond.

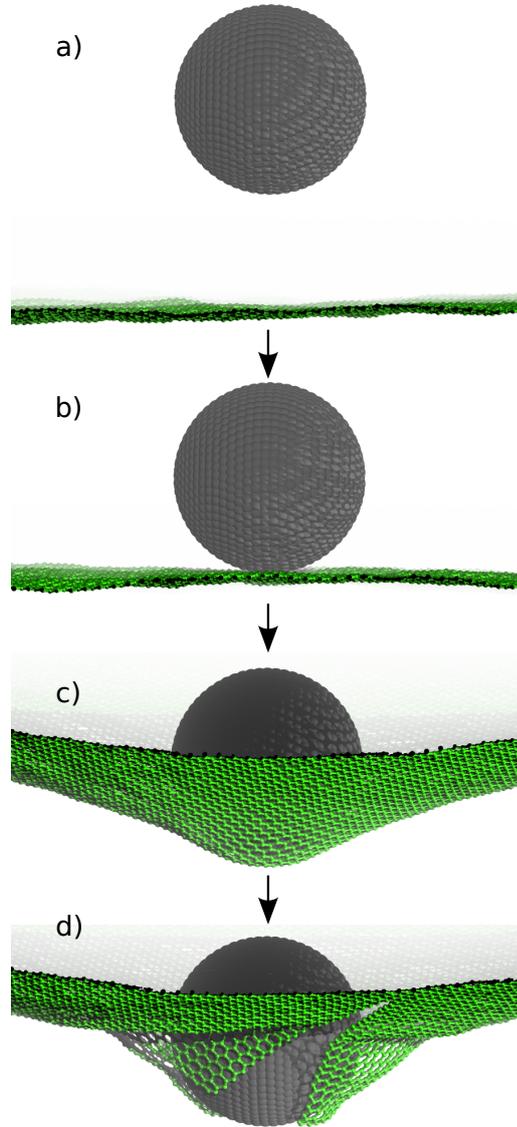
To validate our findings, we investigated the variation of C-C bond lengths in equilibrium MD simulations of small graphene sheets, the amplitude and frequency of ripples formed throughout the sheets due to the thermal fluctuations of the atoms, and elastic parameters (Young's modulus and Poisson's ratio) obtained from in-plane deformations. In all cases, we observed a good agreement between our values and previous experimental or theoretical results<sup>61</sup>.

We also compared the computational efficiency of the truncated Morse potential with that of other potentials which could be used to model graphene, either carbon-specific<sup>56,57</sup> or of a more general nature<sup>54,59</sup>. Our potential was more than 3 times faster than the fastest and more than one order of magnitude faster than the slowest of the other potentials<sup>61</sup>. This result recommends the truncated Morse potential for simulations of mixtures of graphene with biomolecules, as well as for graphene sheets of large dimensions, as described in the next section.

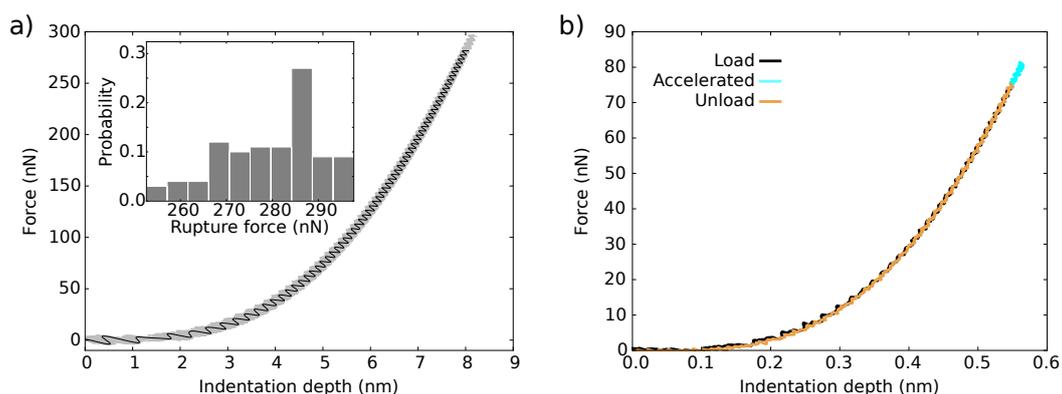
## 3.3 GRAPHENE INDENTATION UNTIL RUPTURE

We performed MD simulations reproducing as close as possible the AFM indentation experiments on graphene<sup>1</sup>. A sphere, mimicking the indenter tip, was moved at constant velocity into a circular sheet which was clamped around the edge. A graphical representation of the different phases is shown in Figure 3.5. All molecular systems are simulated at room temperature (300 K), unlike most other studies which added kinetic energy by slightly displacing C atoms from their equilibrium positions<sup>67,72,74,75</sup>. We used GROMACS<sup>43</sup> 4.5.3 with the truncated Morse potential; for comparison, we also used LAMMPS<sup>76</sup> version 17Feb2012 with the carbon-specific multi-body AIREBO potential<sup>57</sup>. Pairwise forces, per atom punctual stress (Section 2.3.3), and circular stress (Section 2.3.4) were obtained with the TRFDA code (Section 2.2). For calculations of virial atomic stress (Section 2.3.1), we used an own implementation, also based on GROMACS 4.5.3. A detailed protocol can be found in Appendix B.

During the MD simulations, we have recorded load-displacements profiles, representing the dependency of the force on the tip indentation until rupture. Figure 3.6a shows the typical indentation profiles we have obtained from our simulations, in which we observe the same non-linear elasticity of the sheet uncovered by experiments<sup>1</sup>. During the indentation, the force increases in order to maintain the constant velocity of the sphere, indicating a progressive stiffening of the graphene sheet. From a series of 100 independent simulations, started with random velocities, using sphere radii of 5 nm and graphene radii of 25 nm, we calculated an average



**Figure 3.5:** Different phases of the indentation process until rupture. The sphere starts at a certain distance from the graphene sheet (a), comes into contact with the sheet (b) and continues to press on it (c) until rupture (d). The ruptured sheet is included here only for visualization purposes, but the dynamics after bond breaking were not included in our analysis.

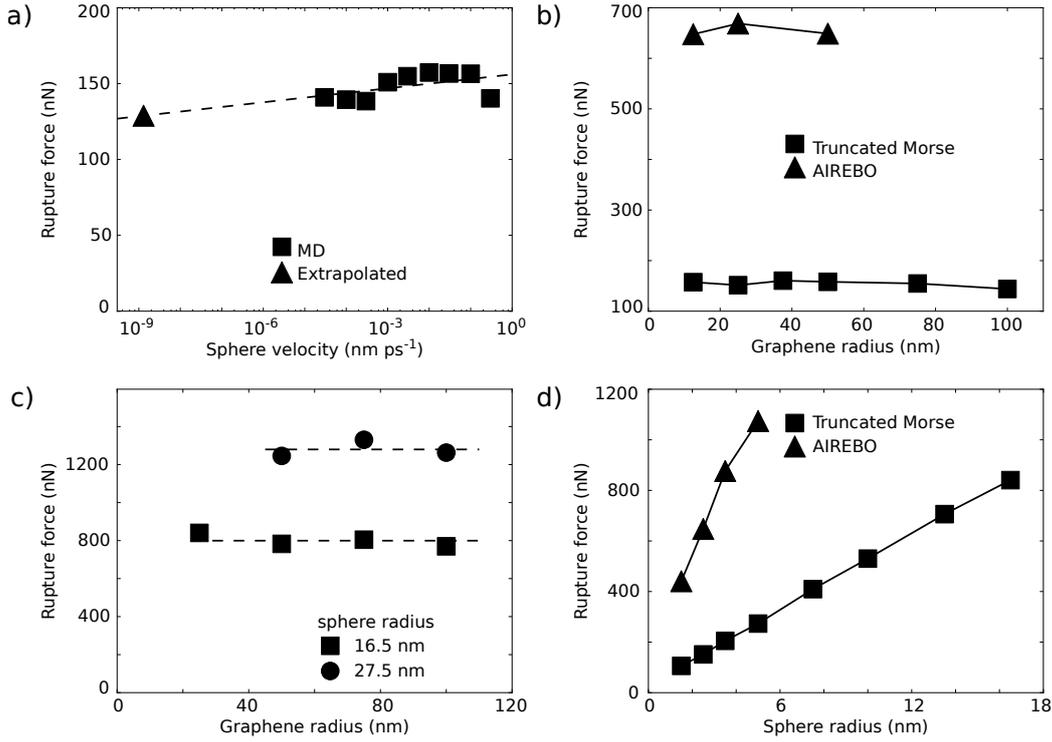


**Figure 3.6:** Graphene force-indentation profiles. a) Overlay of non-linear force-indentation curves for a set of 100 independent simulations with a sphere radius of 5 nm and graphene radius of 25 nm (gray). One such curve, corresponding to the median rupture force, is shown in black. Inset: histogram of rupture forces, with an average of 280 nN. b) Force-indentation profile with reverse load. Initial loading (black) was performed with a constant velocity of  $0.01 \text{ nm ps}^{-1}$ . The accelerated regime (blue) was performed with an initial velocity of  $0.01 \text{ nm ps}^{-1}$  and an acceleration of  $-0.0001 \text{ nm ps}^{-2}$ , for a total time of 200 ps. Unloading (yellow) was performed with a constant velocity of  $-0.01 \text{ nm ps}^{-1}$ . A negative velocity denotes a movement of the sphere in the opposite direction, reducing the load on the graphene sheet.

rupture force of  $280 (\pm 10) \text{ nN}$ . Rupture forces vary among individual trajectories by up to  $\pm 10\%$  (Figure 3.6a inset), the same range as obtained from the AFM experiments<sup>1</sup>. A reverse load simulation showed that the molecular system is close to equilibrium throughout the indentation process prior to rupture (Figure 3.6b).

Similar to the snap-in noticed in experiments (Figure S3a in Reference 1), we observed an initial jump to depth negative values in all the raw indentation profiles derived from our simulations. This is a consequence of attractive Lennard-Jones interactions that occur between atoms of the sphere and atoms of the sheet when they become closer than the cut-off distance. While in experiments the indenter moved towards the sheet, in our MD simulations the sheet moved towards the indenter. The deflection of the sheet had an amplitude comparable to the thermally induced ripples<sup>61,77,78</sup>, such that it did not give rise to any significant change in the geometry of the sheet. The graphene indentation profiles shown in Figure 3.6 and used further in calculations only contain the data from after the tip has reached the initial position of the sheet; this is equivalent to the cantilever becoming straight in an AFM experiment, placing the tip firmly in contact with the sheet.

We next investigated the influence of the indenter velocity on the rupture force. Over a range of velocities covering four orders of magnitude, the rupture force changes very little (Figure 3.7a). By applying the Bell model (Section 1.2), we obtained a linear regression from which we could estimate the rupture force at lower indentation velocities. For the experimental AFM indenter velocity ( $1.3 \cdot 10^{-9} \text{ nm ps}^{-1}$ ), the Bell model predicts a value less than 4% from the lowest rupture force obtained from MD simulations, well within the variability range obtained above. We therefore conclude that the force-indentation profiles and rup-



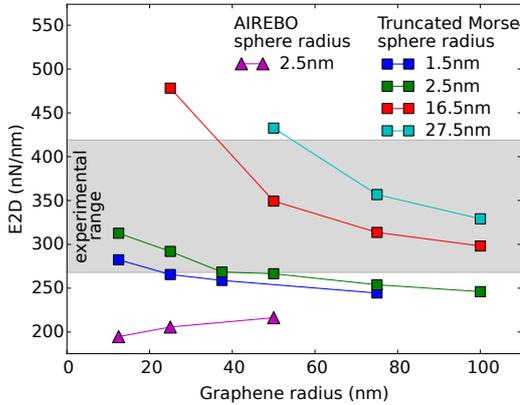
**Figure 3.7:** Variability of rupture force ( $F_R$ ) and the influence of sphere velocity ( $v_s$ ), sphere radius ( $r_s$ ) and graphene sheet radius ( $r_g$ ). a) Influence of  $v_s$  on  $F_R$  over four orders of magnitude (squares) for  $r_s=2.5$  nm and  $r_g=12.5$  nm, using the truncated Morse potential. Lowest  $F_R$  is 133 nN corresponding to  $v_s$  of  $10^{-4}$  nm ps<sup>-1</sup>. For the experimental  $v_s$  of  $1.3e-9$  nm ps<sup>-1</sup> ( $=1.3$   $\mu$ m s<sup>-1</sup>), the  $F_R$  estimated from a linear regression (dashed line) is 129 nN (triangle). b) Influence of  $r_g$  on  $F_R$  using the truncated Morse potential (squares) and the AIREBO potential (triangles) for  $r_s=2.5$  nm and  $v_s=0.01$  nm ps<sup>-1</sup>.  $F_R$  obtained with the AIREBO potential are 4.2 ( $\pm 0.2$ ) times larger than those obtained with the truncated Morse potential. c) Influence of  $r_g$  on  $F_R$  force using the truncated Morse potential for  $r_s=16.5$  nm (squares) and  $r_s=27.5$  nm (circles), with  $v_s=0.01$  nm ps<sup>-1</sup>. Dashed lines represent average  $F_R$  for each sphere radius: for  $r_s=16.5$  nm – 800 nN (experimental 1700 nN), for  $r_s=27.5$  nm – 1280 nN (experimental 2890 nN). d) Influence of  $r_s$  on  $F_R$  using the truncated Morse potential for  $r_g=25$  nm (squares) and the AIREBO potential for  $r_g=12.5$  nm radius (triangles), with  $v_s=0.01$  nm ps<sup>-1</sup>.  $F_R$  obtained with the AIREBO potential is consistently around 4.2 times larger than the value obtained with the truncated Morse potential.

ture forces from our MD simulations can be directly compared with the experimental ones, even though they cover different ranges of indenter velocities. For all further indentation simulations, we used a sphere velocity of  $0.01$  nm ps<sup>-1</sup>.

Lee *et al.*<sup>1</sup> suggested that the rupture force does not depend on the graphene sheet radius, but could not fully test this hypothesis due to experimental limitations. From simulations on a wide range of graphene radii (12.5–100 nm) with both the truncated Morse and the AIREBO potentials, we could confirm this finding (Figure 3.7b). This allows us to use for further comparisons spheres with the same radii as the AFM indenters (16.5 and 27.5 nm) but graphene sheets smaller than the experimental ones (radii of 20–100 nm instead of 1–1.5  $\mu$ m, Figure 3.7c), without changing the outcome. We also observed that the rupture forces obtained with the truncated Morse potential were consistently around 2.2 times

smaller than the experimental values and around 4.2 times smaller than the ones obtained with the AIREBO potential (Figures 3.7b and 3.7d). Predicting a mechanical resistance for graphene of the correct order of magnitude is an excellent result for the truncated Morse potential, which is pairwise potential, much less sophisticated and computationally expensive than AIREBO, and for which no macroscopic material parameters were used.

A dependency of the graphene rupture force on the AFM indenter radius was suggested<sup>1</sup>, but not further characterized, as only two tip radii were available for the experiments. The larger tips (27.5 nm) yielded proportionally higher rupture forces than the smaller ones (16.5 nm), hinting towards a linear relationship. We were able to build spherical models of the indenter tip with a much wider range of radii, between 1.5 and 27.5 nm. In indentation simulations, we observed a perfectly linear dependency of the rupture force on the sphere radius, for graphene sheets using both the truncated Morse potential and the AIREBO potential (Figure 3.7d). While providing a direct extension of the experimental results, this finding is also in excellent agreement with theoretical predictions of a spherical indenter pressing on a circular linearly-elastic plate<sup>2</sup>. This could not be expected, given that graphene has shown a non-linear elastic behavior (Figure 3.6). We conclude that the linear dependency of the rupture force on the tip radius might also hold for the more general case of materials with non-linear elasticity.



**Figure 3.8:** Two-dimensional elastic modulus ( $E^{2D}$ ) for various sphere and graphene sheet radii, using the truncated Morse (squares) and AIREBO (triangles) potentials. The horizontal gray area shows the range of  $E^{2D}$  values obtained from experiments by Lee *et al.*<sup>1</sup>.

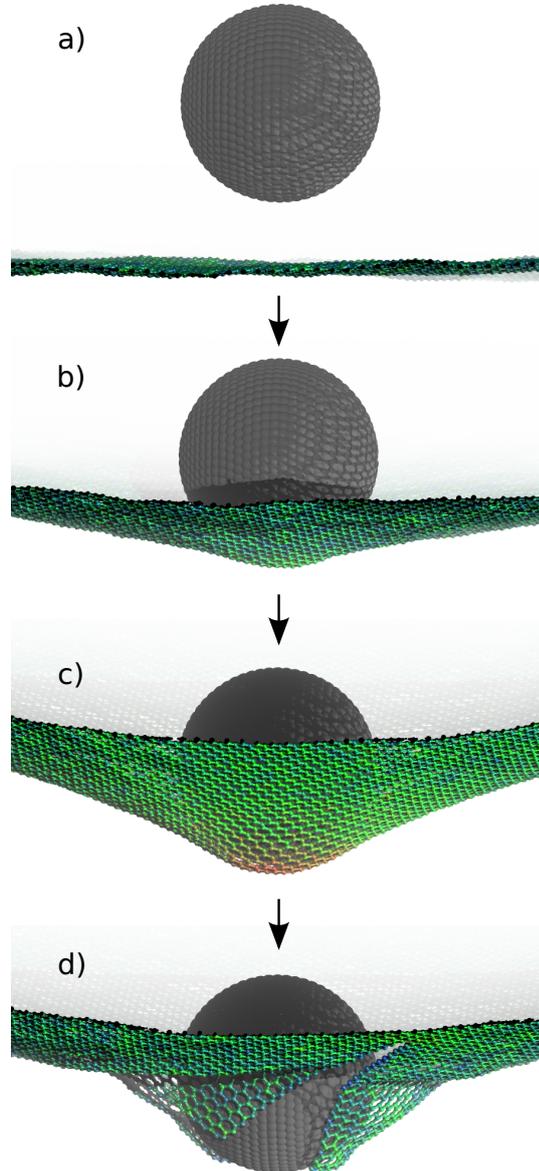
When using the truncated Morse potential, most of the values lay within the  $E^{2D}$  range derived from experimental data. This is an excellent result, offering a further validation of our model. At the same time, the values obtained from simulations with the AIREBO potential were more than 10% smaller than the lower limit of the experimental  $E^{2D}$  range. For each sphere radius, an increase in graphene sheet radius had the effect of decreasing  $E^{2D}$ ; we believe that this is the result of a too low ratio between the sheet and sphere radii, which allows edge effects to influence the appar-

From AFM load-indentation profiles, Lee *et al.*<sup>1</sup> have computed a two-dimensional elastic modulus ( $E^{2D}$ ).  $E^{2D}$  differs from the typical elastic (Young's) modulus, as it avoids defining the thickness of the graphene sheet, which is still controversial<sup>79</sup>. For this reason, it is often used for comparison between experimental and theoretical models of graphene<sup>68-70</sup>. We have calculated  $E^{2D}$  values for various sheet and tip radii (Figure 3.8) from load-indentation profiles obtained in our MD simulations.

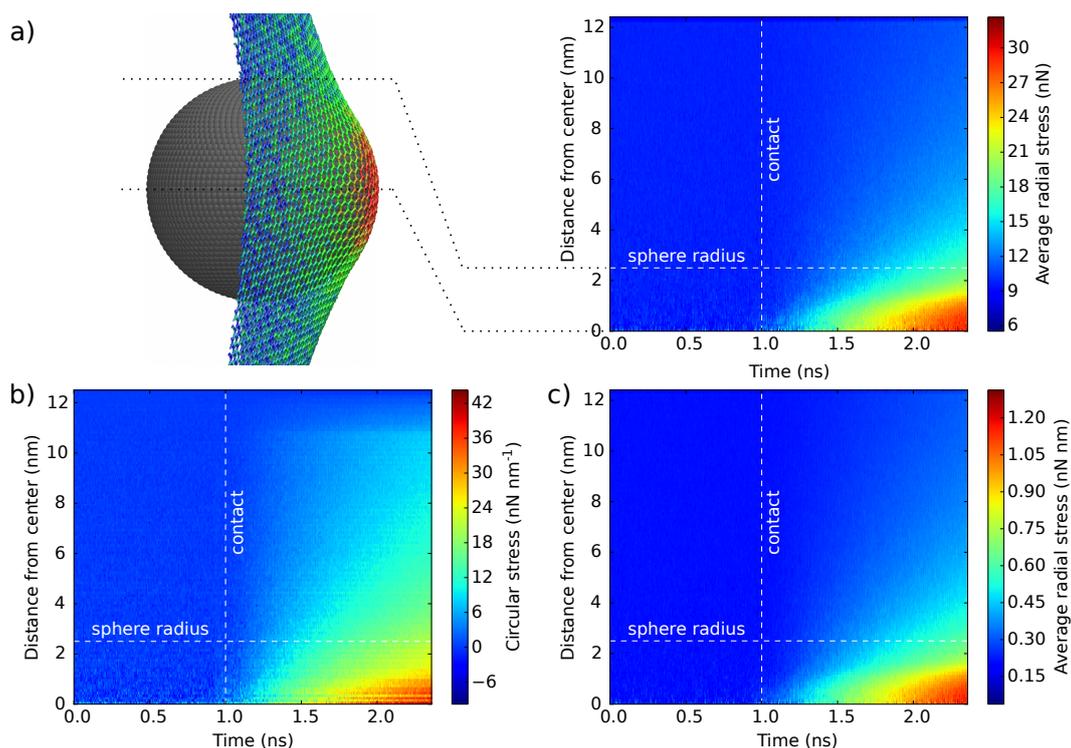
ent strength of the sheet. Indeed, the  $E^{2D}$  values appear to converge as the sheet radius increases, and small radii spheres show a faster convergence than larger ones, supporting our hypothesis.

Having validated our computational model, we next aimed at explaining the material failure as well as its dependence on sheet and tip radii by following the variation in time of the internal stress during the graphene indentation until rupture. The atomic level stress does not have a unique definition; in Section 2.3, we explored two existing definitions and introduced two new ones, based on the calculation of pairwise forces in TRFDA. To prevent a bias in our results, we used three of the four stress definitions: punctual stress (Section 2.3.3), circular stress (Section 2.3.4), and virial stress (Section 2.3.1). The local pressure (Section 2.3.2) relies on rectangular grid cells, and therefore does not fit well the circular symmetry of our system. We note again that none of these three expressions is a true stress, using units of force, force over length, and force times length (or energy), respectively. Normalization by the area of action of pairwise atomic forces, the height of the single atom thick layer, or the atomic volume, respectively, would yield true stresses, however these quantities are ill-defined and, thus, not included in calculations. We can consider though that they are constant throughout the sheet, as graphene is a homogeneous material.

The distribution of stress in the graphene sheet at several different moments during the indentation process is shown in Figure 3.9. The indentation process starts with



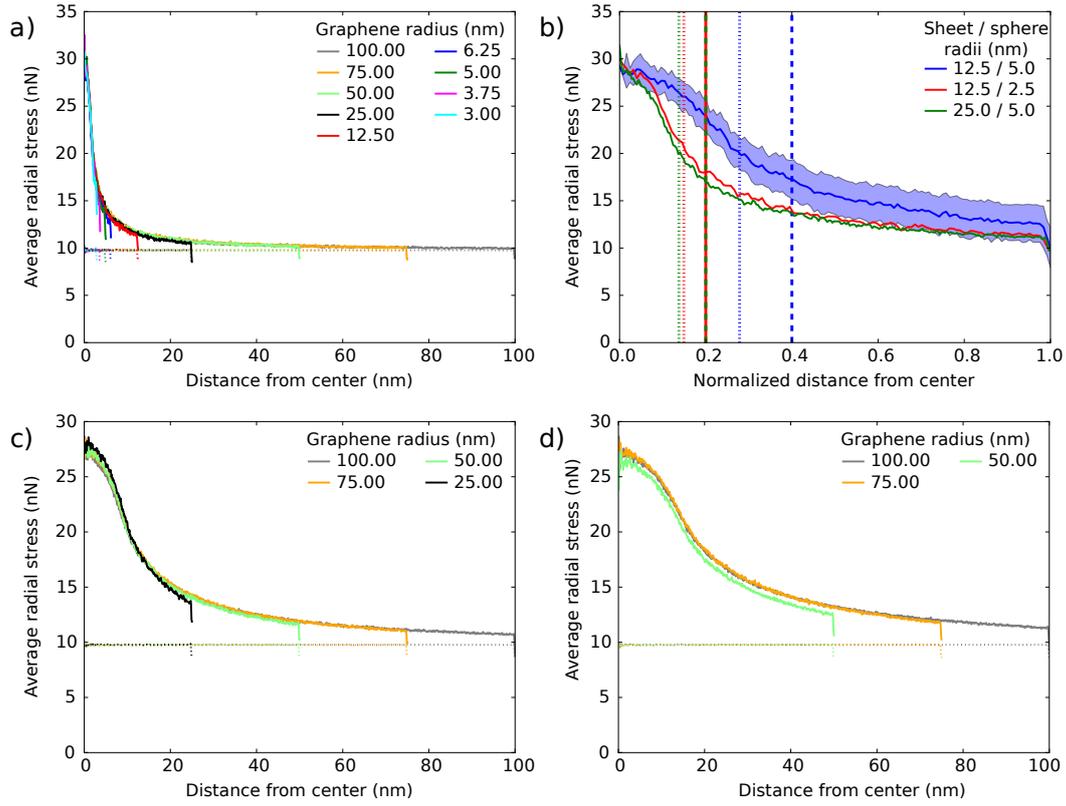
**Figure 3.9:** Phases of the indentation process until rupture, with the sheet atoms colored by punctual stress. a) While the sphere is away from the sheet, thermal ripples generate a low level of background stress. b) At the start of indentation, the stress under the indenter remains low. c) Just before rupture, stress is concentrated under the indenter. d) Stress returns to low levels immediately after rupture. The ruptured sheet is included here only for visualization purposes, but the dynamics after bond breaking were not included in our analysis.



**Figure 3.10:** Stress variation during indentation and just before rupture, for a graphene radius of 12.5 nm, a sphere radius of 2.5 nm, and velocity of  $0.01 \text{ nm ps}^{-1}$ . The radial stress was obtained by averaging the per atom punctual stress (a) or virial atomic stress (c) over all atoms found at the same distance from the center of the graphene sheet. The circular stress (b) was defined along concentric circles and did not include averaging. The distance from the center was computed with a resolution of 0.1 nm. The snapshot in a) shows the distribution of per atom punctual stress, 1 ps before the rupture, in a side view.

the sphere at a certain distance from the sheet; this is required for the sphere to reach and maintain the desired velocity. Also from the start of the simulation, spontaneous ripples appear throughout the graphene sheet due to thermal fluctuations<sup>61,77,78</sup>; the relative movement of the atoms with respect to each other causes a low level of background stress. The snap-in which occurs when the sphere and sheet become closer than the cut-off distance, as well as any possible shock waves formed due to the contact, cannot be distinguished from the background stress, suggesting a high flexibility of the sheet. As the indentation continues, the tip presses onto the sheet; the stress increases significantly only in and around the area of graphene in contact with the sphere, while it remains at levels comparable to the background stress towards the edge of the sheet. Timelines of the stress distribution during the indentation process are represented in Figure 3.10 separately for the three types of atomic stress. We note that they are very similar, even though they are expressed in different units and can be compared only qualitatively.

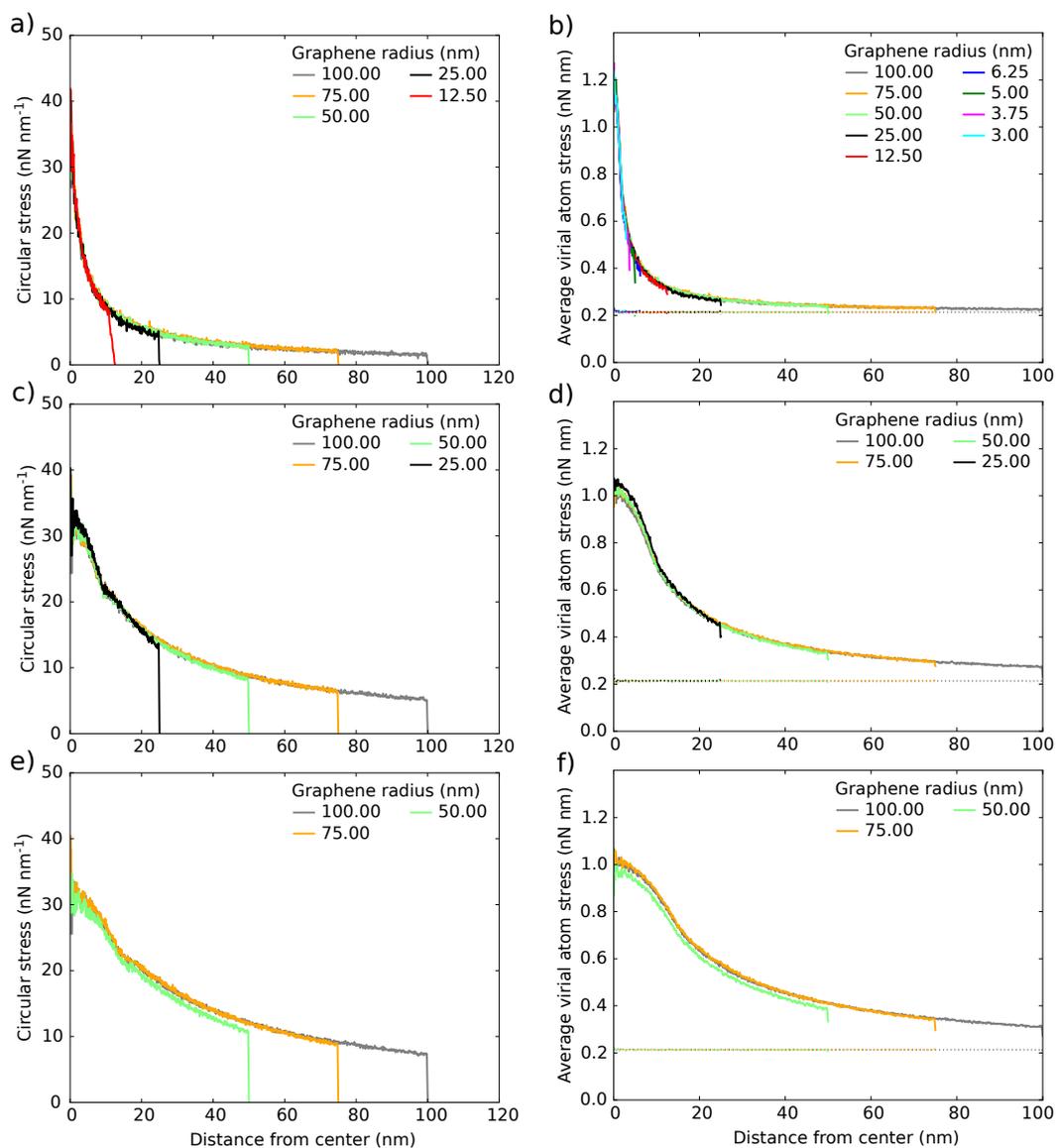
We next investigated the correlation between the stress distribution, measured shortly before rupture, and the sheet and tip radii. We found that the radial stress at a certain



**Figure 3.11:** Distribution of per atom punctual stress less than 100 ps before rupture for several graphene sheet radii ( $r_g$ ) and sphere radii ( $r_s$ ). The radial stress was obtained by averaging the per atom punctual stress over all atoms found at the same distance from the center of the graphene sheet. a,c,d) Radial stress as a function of distance from the center of the sheet for  $r_s=2.5$  nm (a),  $r_s=16.5$  nm (c),  $r_s=27.5$  nm (d) and several  $r_g$ . Dotted lines show the background stress. b) Radial stress for several  $r_g/r_s$  pairs: 12.5/2.5 (red), 12.5/5.0 (blue) and 25.0/5.0 (green), as a function of distance from the center normalized by  $r_g$ . The shaded area shows the standard deviation for one of the curves. Dashed lines show  $r_s$ , dotted lines show the radii of the contact area between the sheet and the sphere.

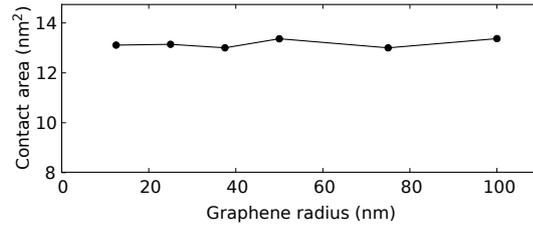
distance from the center of the sheet has no dependency on the sheet radius (Figure 3.11a); this relationship holds even for sheet radii approaching the sphere radius (3 nm versus 2.5 nm). High stresses are strongly localized in the area of graphene under the sphere (also visible in Figure 3.10); the stress near the edge of the sheet reaches background levels for a ratio of sheet to sphere radii of 5–10 (Figures 3.11b,c,d), significantly smaller than assumed by the theoretical models<sup>1,2</sup>. The stress at the center of the sheet is higher than at the edge of contact between the sphere and the sheet (dotted lines in Figure 3.11b), with a ratio between 1.4 and 1.9, while the theoretical model for a linear elastic material<sup>2</sup> suggests a lower value of 1.17. We also note that the value of critical radial stress of  $\sim 30$  nN at the center of the stress, at which rupture occurs, remains the same independent of the sheet and sphere radii. The analysis of circular stress and virial atomic stress revealed again very similar distributions (Figure 3.12).

In a related result, the area of contact between the sphere and graphene was independent of the sheet size (Figure 3.13). Together with the stress distribution findings, this points towards a small area under and around the indenter, determined only by the indenter radius, which is sufficient to bear the load by distributing the stress among the constituent C-C bonds. This central area is mechanically supported by the rest of the sheet, throughout which the stress has much lower levels.



**Figure 3.12:** Distribution of circular (a, c, e) and radial virial (b, d, f) stress less than 100 ps before bond breaking for a sphere radius of 2.5 nm (a, b), 16.5 nm (c, d) and 27.5 nm (e, f) and several graphene sheet radii. The radial virial stress was obtained by averaging the virial atomic stress over all atoms found at the same distance from the center of the graphene sheet. The distance from the center was computed with a resolution of 0.1 nm. Dotted lines show the background stress.

Figure 3.11b also shows the strong influence the sphere radius has on the stress distribution. When the sphere radius is doubled from 2.5 to 5 nm while the sheet radius is kept constant, the contact area between the sphere and graphene increases and the stress distribution becomes wider. In agreement, we also found that reaching the critical radial stress requires a higher indentation force ( $\sim 284$  nN) for the larger sphere compared to the smaller one ( $\sim 160$  nN).



**Figure 3.13:** Contact area between a sphere of 2.5 nm radius and graphene sheets of various radii less than 100 ps before rupture. Atoms are considered in contact if the distance between them is less or equal to 0.4 nm.

While the stress distribution depends on the indenter radius and does not depend on the sheet radius, stress profiles along normalized distances from the center overlap for the same ratio of sheet and sphere radii (Figure 3.11b). Thus, it is this ratio that determines the behavior of the entire sheet during indentation, while the rupture force depends solely on the sphere indenter radius.

To investigate this link in more detail, we turned to study how the spatial distribution of stress over the C-C bonds determines their likelihood to reach the critical point. From 100 independent MD simulations using the same parameters except for random initial velocities, we calculated the probability of bonds to break as a function of their distance from the center of the sheet (Figure 3.14a). We observed a very fast decrease of the breaking probability as the distance increased, with the probabilities becoming lower than 1% at distances larger than 1.2 nm, which are well within the sphere radius. This agrees very well with the experimental finding that graphene rupture starts only under the indenter<sup>1</sup> and reflects the high concentration of stress under the AFM tip.

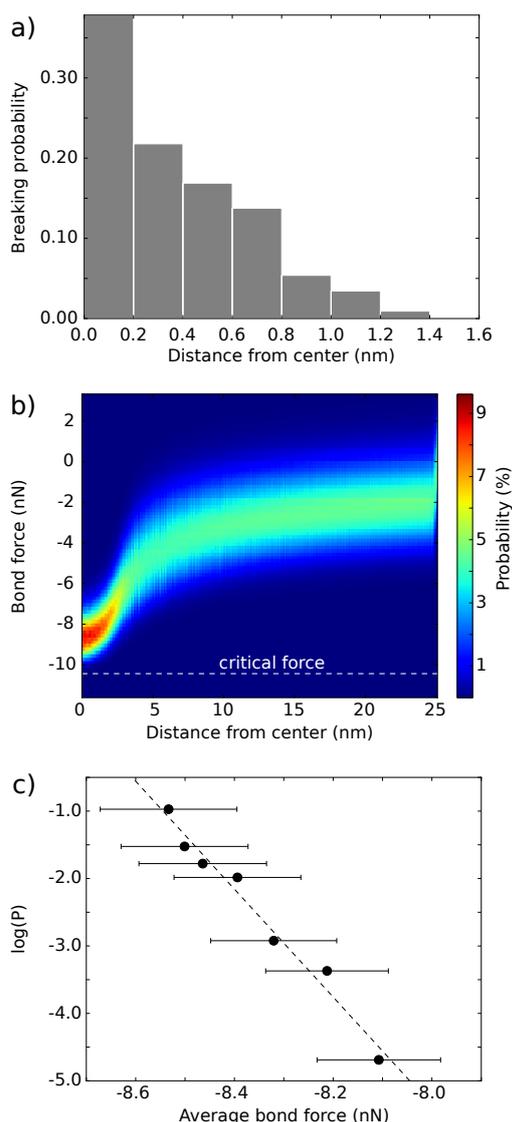
Given the rapid decay of bond rupture probabilities on the sub-nanometer scale, we investigated the influence of the local geometry of the network of C-C bonds on our findings. The C-C bonds which are closest to the center of the sheet are the most likely to break (Figure 3.14a), therefore changes in their spatial arrangement would have the largest impact on the outcome. From the point of view of the contact between the lowest atom of the sphere and the sheet, there are two extreme cases. If an atom is placed at the center of the graphene sheet, the indentation leads to stretching the three C-C bonds formed by this atom, which is the case in all our indentation simulations performed so far. In contrast, if the center of the sheet coincides with the middle of an aromatic ring, the indentation leads to stretching of the six C-C bonds forming the ring. From a small set of simulations comparing these two cases, we obtained results indistinguishable from each other. This suggests that a small displacement of the sphere with respect to the center of the sheet does not bear any influence on the results, and is consistent with the finding that bonds

**Figure 3.14:** Probabilities of bond breaking from a series of 100 independent simulations with a sphere radius of 5 nm, a graphene sheet radius of 25 nm and a sphere velocity of  $0.01 \text{ nm ps}^{-1}$ . The distance was calculated between the center of the sheet and the closest of the two atoms forming the bond which breaks, with a resolution of 0.1 nm. By convention, an attractive force is negative.

a) Probability of bond breaking as a function of distance from the center of the graphene sheet. The number of broken bonds in each bin was normalized by the total number of bonds in that bin.

b) Distribution of C-C bond forces as a function of distance from the center of the graphene sheet. For each distance from the center, the distribution reflects C-C bond forces from the last 100 ps before the first bond breaking.

c) Probability of bond breaking as a function of the bond force. The averages and standard deviations are calculated over data from the 100 independent simulations; each data was the median of the forces in bonds located at the same distance from the center of the sheet for the last 100 ps before the first bond breaking. Error bars indicate the standard deviation. The dashed line represents a Bell model fit with  $\Delta x_{\text{bond}} = 0.034 \text{ nm}$  and an average bond force at 100% probability of  $-8.7 \text{ nN}$ .



other than the three formed by the central atom can also break (Figure 3.14a). The AFM indenter in experiments<sup>1</sup> could not be positioned exactly in the middle of the sheet, and had an error of up to 50 nm, further supporting our results. We note that our model of the graphene sheet did not assume a particular bond or bonds in certain regions of the sheet to break, in contrast to previously investigated models based on quantum mechanics or hybrid quantum mechanics/molecular mechanics<sup>80,81</sup>.

Our truncated Morse potential defines the bond breaking upon reaching the critical force of  $\sim 10.4 \text{ nN}$ . Having determined the probabilities of bond breaking as a function of distance from the center of the sheet, we next link the probability of reaching the critical force to the distribution of forces in the individual C-C bonds, as obtained from TRFDA (Figure 3.14b). The forces have large variations but small average values in the outer regions of the sheet, indicating a low probability of bond breaking. In contrast, forces are significantly larger

and the statistical distribution narrower at small distances from the center of the sheet, and in particular within the sphere radius (5 nm), implying a significantly higher probability of bond breaking. Indeed, the probability has an exponential dependency on the average tensile force in the bond (Figure 3.14c). While the determined probabilities only extend to around 10%, by extrapolation to a probability of 100% we find a force of  $\sim 8.7$  nN, slightly lower than the critical force of 10.4 nN. This suggests that a decisive role in propagating a tensed bond to the broken state is played by the thermally induced fluctuations of bond forces.

The graphene rupture can be interpreted as a two-state kinetic process, in which the extension of a single bond beyond the critical point defined by the truncated Morse potential drives the transition between states. Regarding the bond length as the reaction coordinate, we applied the Bell model (Section 1.2) to the data in Figure 3.14c and obtained a distance  $\Delta x_{\text{bond}} = 0.034$  nm between the reactant and the transition state. This is comparable to the difference between the average bond length (0.142 nm) in equilibrium simulations and the critical bond length (0.184 nm) for the truncated Morse potential. Regarding the indentation depth as the reaction coordinate, the Bell model could also be applied to the rupture force dependency on the indenter velocity (Figure 3.7a), leading to a distance  $\Delta x_{\text{indent}} = 1.4 \cdot 10^{-3}$  nm. The discrepancy of more than one order of magnitude between  $\Delta x_{\text{bond}}$  and  $\Delta x_{\text{indent}}$  probably originates in the different orientation of the reaction coordinate. The surface of the sphere has a low local curvature on the scale of the C-C bond length, such that the bonds under the indenter are nearly perpendicular to the indentation direction; consequently, projecting the elongation of the bonds on this direction results in a very small  $\Delta x_{\text{indent}}$ . Thus, while the energy landscape of the single bonds (*i.e.* the shape of the truncated Morse potential) determines the kinetics of individual bond breaking, the projection of that energy landscape along the direction of the indentation governs the rupture of the graphene sheet.

### 3.4 DISCUSSION AND CONCLUSIONS

We performed MD simulations replicating as closely as possible the AFM experiments of graphene indentation<sup>1</sup>. Instead of deriving elastic parameters exclusively from deformation studies<sup>68,69,79</sup>, we focused on the material rupture and the internal stress distribution as its molecular basis. We measured load-indentation profiles and determined rupture forces for graphene sheets free of defects. The rupture forces showed the same variability as in experiments and a very weak dependency on the indentation velocity, allowing us to make direct comparisons between our results and those of Lee *et al.*

In order to carry out these studies for sheets reaching almost experimental dimensions, we developed the truncated Morse potential, a pairwise potential able to represent C-C bond breaking in graphene. In MD simulations using this potential, the deformation and rupture results are as close to the experimental ones as those obtained with a much more complex bond order potential (AIREBO). The model using the truncated Morse potential also predicted the elasticity of the sheet, expressed through the two-dimensional elastic modulus ( $E^{2D}$ ), better than AIREBO. Unlike the bond order potentials, the truncated Morse potential is not suitable to represent bond reformation between the same or other atoms. However, this does not represent a disadvantage in mixed simulations involving much softer biomolecules, where the C-C bonds in graphene are not stretched close to their critical point. Furthermore, our potential is much simpler to implement and more computationally efficient than other models used for carbon-based compounds. These strengths recommend the truncated Morse potential for further mechanical studies of graphene, especially in mixed graphene-biomolecular simulations or for carbon-based molecular systems involving millions of atoms in which bond reforming is not expected.

We used the truncated Morse potential to perform indentation simulations until rupture, from which we studied the relationship between the rupture force and some geometrical parameters as well as the stress distribution during the process. In excellent agreement with results from AFM experiments<sup>1</sup>, we found the rupture force to be independent of the sheet radius and to linearly depend on the indenter radius. The latter finding was surprising, as the linear dependency on the tip size was theoretically predicted for a linearly elastic material<sup>2</sup>, while graphene showed a non-linear elastic behavior.

The atomic stress distribution just before graphene rupture suggests that a small area of the sheet found under the indenter is able to bear most of the load. The rest of the sheet offers mechanical support to this central area and can, in fact, be very small; the stress distribution and the rupture force remain almost unchanged down to a ratio of sheet to sphere radii close to one. This is an unexpected finding, as it was previously assumed that this ratio needs to be much larger for consistent measurements of mechanical parameters. Therefore, we suggest that the outcome of similar AFM experiments would not change if they are performed with larger indenters or smaller graphene sheets. This result is of high practical importance, as obtaining smaller defect-free graphene sheets is easier and associated with reduced production costs.

The stress distribution just before rupture differs from the ones predicted by an FE model<sup>1</sup> and an analytical model for elastic materials<sup>2</sup>. This contrasts to the previous finding of the linear dependency of rupture force on the indenter radius, which fits both graphene and linear elastic materials. Developing a generalization of current models to describe non-

linearly elastic materials like graphene would therefore be desired, but is beyond the scope of this work.

Our analysis of tensile forces in highly stretched C-C bonds suggests that thermal fluctuations initiate the material rupture. The probability of a bond to break decays very fast as the distance from the center of the sheet increases, such that material rupture is restricted to a small area under the indenter for all practical purposes, in excellent agreement with experimental observations<sup>1</sup>. Although the truncated Morse potential was parametrized for C-C bonds in graphene, our analysis also holds for other materials for which such a potential is an appropriate model. For example, if the dissociation energy remains the same, a broader Morse potential implies a larger critical bond length and a lower critical force; thermal fluctuations in the indented material would be able to more easily stretch the bonds to their critical point, resulting in an increased area around the center of the sheet with high probabilities of bond breaking. Conversely, a narrower Morse potential with the same dissociation energy would be associated with a higher localization of the rupture initiation point under the AFM indenter.

Two geometrical coordinates drive the graphene rupture at different scales: the macroscopic indentation of the sheet leads to stress distributions and rupture forces dependent on the indenter radius, while the microscopic stretching of the C-C bonds determines the location of the rupture initiation. Although these results have been obtained for defect-free graphene sheets, we can also use them to predict the rupture behavior of sheets with defects under an indentation load. Few, small defects can potentially generate a localized change in the stress distribution and an associated increase in the range of bond forces in that area. If such defects are found under the indenter, where the bond forces have large values and a narrow statistical distribution, the probability of reaching the critical force would become higher, leading to a smaller rupture force. If the defects are located in the outer regions of the sheet, their impact on the already wide range of low forces would be minimal and the rupture force would remain virtually the same. However, defects covering wide areas or in large numbers will significantly disturb the regular network of C-C bonds, such that the stress distribution as well as the probability and location of the sheet rupture can no longer be inferred from the current results.

Our findings regarding the very weak dependency of the rupture force on the indenter velocity suggest that the load distributes very quickly and efficiently among the network of C-C bonds, generating a radially uniform stress concentration. If the dissipation of stress away from the point of application would be slow, it would lead to localized areas with increased stress levels and higher bond force. Similar to the case of defects, this would translate into higher bond breaking probabilities and lower sheet rupture forces. Although possibly the result of insufficient statistical sampling, the slightly lower rupture force ob-

tained at the higher end of the scale in Figure 3.7a could also be an indication of the indentation velocity becoming close to the load dissipation rate, leading to such an uneven stress distribution. Analyzing such fast processes is possible with TRFDA, but requires a much finer resolution of the time-resolved pairwise forces data, and will be the subject of future work.

# 4

---

## MECHANICAL UNFOLDING OF PROTEINS

---

In this chapter, we investigate the unfolding of proteins under the influence of an external force, probed through force-clamp MD simulations. We initially focus on the kinetic aspects of the forced unfolding, comparing the kinetic profiles with experimental ones. We then use TRFDA to investigate how the applied force distributes through the protein structure, and whether this distribution has an influence on the unfolding kinetics.

### 4.1 INTRODUCTION

The development of techniques through which proteins can be unfolded mechanically has enabled measuring the response of proteins to an externally applied force. Single molecule force spectroscopy (Section 1.1) or molecular tweezers have been used to mimic the mechanical stresses that biomolecules experience in their native environment. Among others, these techniques can be used to study the relationship between the structure and function of mechanically active proteins (*e.g.* the muscle protein titin), or the transduction of

mechanical signals into the cell and downstream signaling pathways by mechanosensory proteins (e.g. ankyrin repeat proteins, involved in sensing pressure changes in the inner ear)<sup>9</sup>.

Studies of mechanical unfolding of proteins are also relevant in the context of structural biology. The mechanical properties, in particular the resistance to unfolding, are necessarily related to the protein structural properties like topology, patterns or networks of hydrogen bonds (H-bonds) and salt bridges, or core interactions holding together elements of secondary structure. The existing data suggests that  $\beta$ -sheet proteins are the most resistant to unfolding, owing to the H-bonds formed by the neighboring  $\beta$ -strands, but the relative topology and geometry of these strands play an important role.

Ubiquitin was shown to be mechanically resistant, with the force needed to unfold it depending on the orientation of the force and the points where the force is applied<sup>11,82,83</sup>. As ubiquitin appears naturally in multi-unit chains, it can be easily used in force spectroscopy experiments (Section 1.1). The unfolding of individual domains was shown to be independent of each other, and the fraction of unfolded ubiquitin units to have an approximately single exponential time dependence<sup>13</sup>. Single exponential kinetics are in line with a two-state unfolding process, from a well-defined folded to a well-defined unfolded state, in absence of intermediates or other transient states. However, later experiments<sup>3</sup> and analysis<sup>84</sup> have revealed that a stretched exponential (also known as Weibull) function is the best fit for the unfolding kinetics of ubiquitin. Further experiments<sup>4</sup> have replicated this behavior also at single monomer level and for other protein domains (I27 domain of titin, with a  $\beta$ -only secondary structure), suggesting that the stretched exponential kinetics is not characteristic to ubiquitin, but can be found more generally in the mechanical unfolding of proteins. Several explanations have been proposed, among them the fluctuations in the applied force<sup>85</sup> or a glass-like restriction on possible conformations<sup>3</sup>, leading to a model of static disorder in the transition states<sup>5</sup>.

This universal character of stretched exponential kinetics seemed contradicted by results of force spectroscopy experiments on NuG2, a redesigned mutant of the B1 domain of protein G<sup>86</sup>. Cao *et al.* showed that the unfolding of individual domains is independent of each other, similar to poly-ubiquitin, but concluded that NuG2 has single exponential unfolding kinetics<sup>14</sup>. This is even more surprising as NuG2 shares a similar  $\beta$ -grasp fold with ubiquitin. Later analysis<sup>84</sup> suggested though that this contrasting outcome might be due to an alternative way of interpreting the experimental results, in particular binning of the data possibly hiding the differences between a single and a stretched exponential kinetic profile.

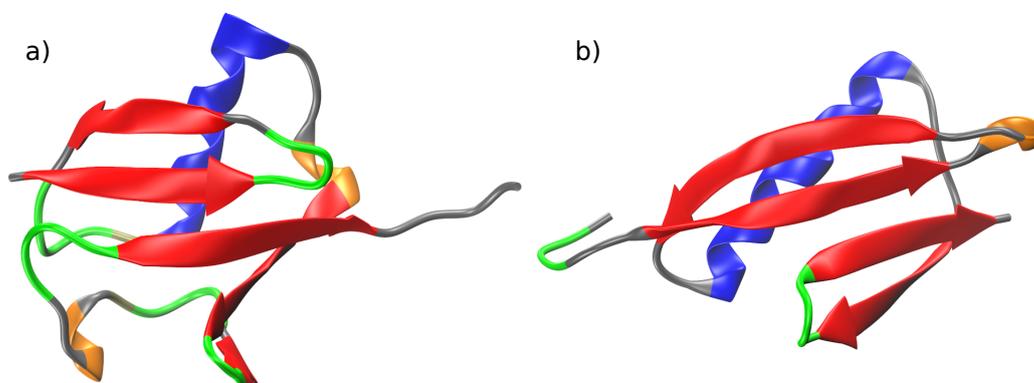
We here set out to reproduce the mechanical unfolding of the two proteins through force-clamp MD simulations and study their unfolding kinetics. To characterize the unfolding

process, we consider both the unfolding times, similar to the experiments, and the distribution of forces or stresses in the three-dimensional structure of the proteins, as obtained from TRFDA. We attempt to reconcile the contradicting results on the shape of the unfolding kinetic profile obtained from experiments. We also study the response of the proteins when forces of different magnitudes are applied, and investigate the validity of the Bell and Dudko-Hummer-Szabo models (Section 1.2) in deriving unfolding rate constants and distances to transition states from mechanical unfolding results. We make extensive use of TRFDA in order to characterize the stress distribution in the stretched proteins, to elucidate their mechanism of forced unfolding and to suggest an explanation for the stretched exponential kinetics. The results obtained here are likely to be highly relevant in the context of further studies of protein kinetics.

#### 4.2 STRUCTURAL ASPECTS OF PROTEIN UNFOLDING

Ubiquitin is a protein with 76 aminoacids, organized in 5  $\beta$ -strands, one  $\alpha$ -helix and a short  $3_{10}$ -helix<sup>87</sup> (Figure 4.1a). The last four aminoacids do not form any secondary structure, constituting a flexible tail. The  $\beta$ -strands are arranged in a  $\beta$ -sheet, with  $\beta 1$  and  $\beta 5$  oriented in parallel and the other three anti-parallel. According to the experimental evidence<sup>87</sup>, most of the aminoacids are part of secondary structure forming H-bonds. The characteristic  $\beta$ -grasp fold is called the “ubiquitin fold” and is found in other proteins, like the SUMO (Small Ubiquitin-related MOdifier) family or ISG15 (Interferon Stimulated Gene).

Ubiquitin is found in all eukaryotic organisms, where it is highly conserved. It has several roles, the main one being to mark other proteins for destruction. For this function, it attaches through an isopeptide bond of its Gly76 residue to Lys residues in its targets. Poly-ubiquitin chains can be formed either by connecting monomers through their N-C ter-



**Figure 4.1:** Structure of ubiquitin (PDB code 1UBQ<sup>87</sup>) (a) and NuG2 (PDB code 1MIo<sup>86</sup>) (b). Secondary structure elements are color-coded:  $\alpha$ -helices in red,  $\beta$ -strands in blue,  $3_{10}$ -helices in orange, turns in green.

mini or by attaching Gly76 of one monomer to one of the 7 Lys residues found in another monomer; both linear and branched chains are possible, the linkage type being related to the role ubiquitin performs in the cell<sup>88,89</sup>. A chain of at least 4 ubiquitin units linked through Lys48-Gly76 isopeptide bonds and attached to a protein marks the protein for degradation in the proteasome. Although ubiquitin was found to be mechanically stable, it is unclear whether this stability has any significance for its function in the cells.

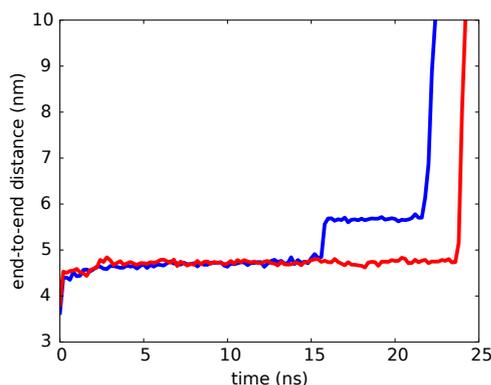
NuG2 is a mutant of the B1 domain of protein G, designed in a computer-aided process to fold around 100 times faster than the native structure<sup>86,90</sup>. It shares a similar  $\beta$ -grasp fold with ubiquitin, with one  $\beta$ -strand less than ubiquitin (Figure 4.1b). The first aminoacids form a flexible region, without any secondary structure. We note that the length of the polypeptide chain used throughout this work (62 aminoacids) corresponds to chain B of the Protein Data Bank structure 1Mlo<sup>86</sup>, which includes a short expression tag; due to the existence of the flexible N-terminal region, the additional aminoacids do not influence the results, except for a slight increase in the end-to-end distance.

In the native state, both proteins have an approximately globular shape. When an external force is applied, they become elongated up to the maximum length of the polypeptide chain. The difference in length between the native and fully stretched states has been measured in force spectroscopy experiments, and is  $20 \pm 0.9$  nm for ubiquitin<sup>3,5</sup> and around 15 nm for NuG2<sup>14</sup>. Under a constant external force, the elongation is observed as a single step transition in most experiments, corresponding to a simple Markovian two-state process. These results have led to the initial conclusion that the unfolding kinetics has a single exponential time dependence. However, unfolding of ubiquitin occurred sometimes through intermediate states, even at different forces<sup>13</sup>; for most of these states, the characteristic elongation was around 8.1 nm. Such intermediate states have also been observed in Monte Carlo simulations using a simplified potential in vacuum<sup>82</sup>. The existence of such intermediates is of particular importance for the kinetic studies, as they might represent relatively stable states along the forced unfolding trajectories and, thus, could lead to a deviation from the two-state character of the process.

To study the mechanical unfolding of proteins, we used force-clamp MD simulations, in which the setup tries to replicate closely the experimental conditions. In short, after solvation and a short equilibrium simulation, an external force was applied on the C $\alpha$  atoms of the N- and C-termini. The distance between these atoms represents the end-to-end distance, which was recorded and represented the criterion used to determine the evolution of the unfolding process. The simulations were stopped when an end-to-end distance of 10 nm (for ubiquitin) and 8 nm (for NuG2) has been reached, allowing the proteins to elongate by at least 4–5 nm from the native structure. Stopping the simulations before reaching the maximum polypeptide chain length is motivated by the sudden character of

the elongation observed in experiments, which suggests that, once started, the unfolding proceeds quickly until the full stretching of the protein. Even in the few cases where an intermediate state has been observed for ubiquitin, the elongation between the native state and the intermediate, as well as between the intermediate and the fully stretched state, is sudden. Therefore, apart from a small decrease in the overall unfolding time, this choice should not bear an influence on our results. A more detailed description of our simulation setup can be found in Appendix C.

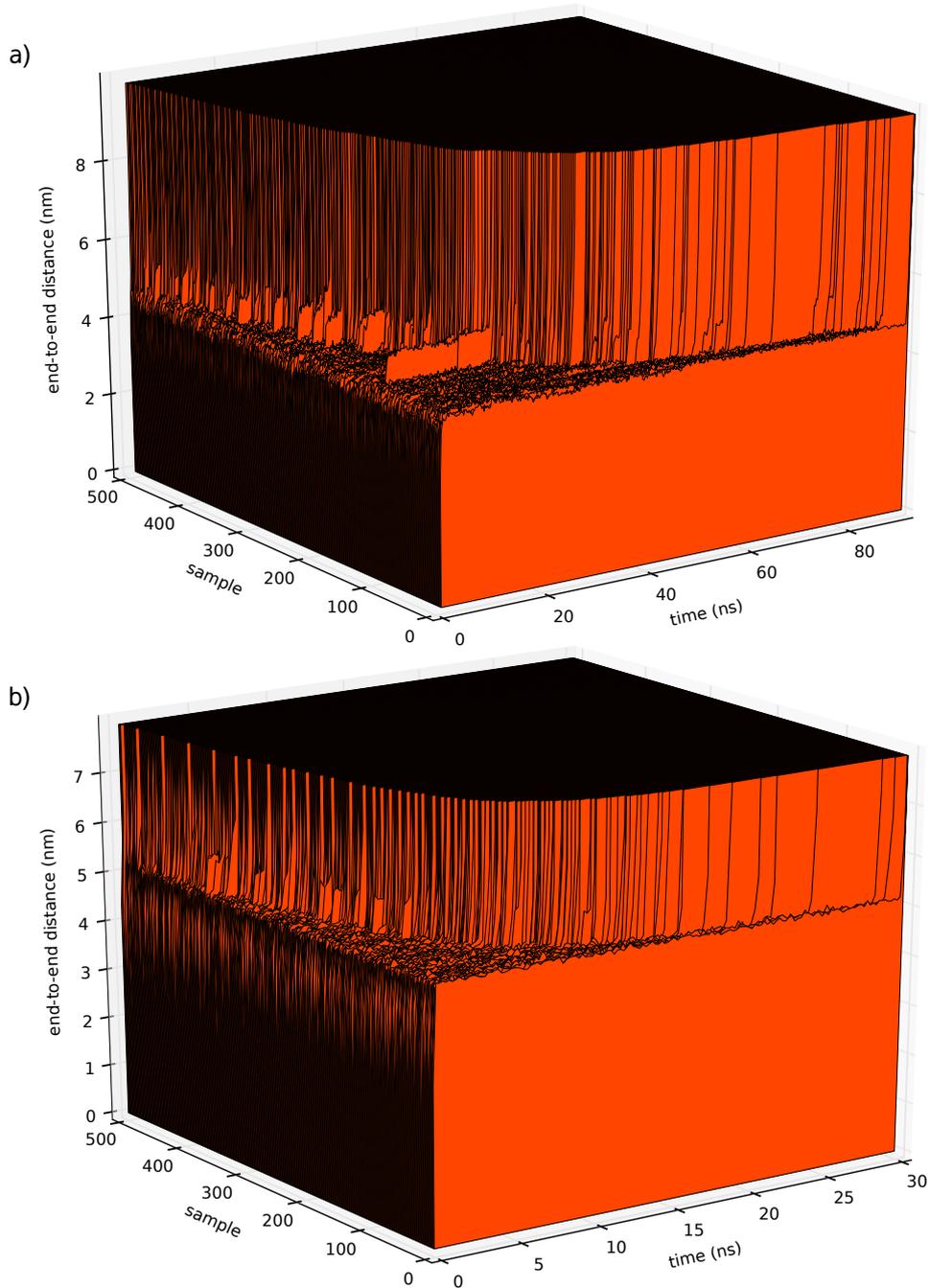
Although we apply several different forces on each protein, in most cases the evolution of the end-to-end distance during the MD simulation has the same three phases: an initial rapid increase, a waiting time, and a final rapid increase (Figure 4.2). The initial increase starts from the native end-to-end distance and ends with reaching the plateau characteristic for the waiting time; this phase is short, taking in most cases below 2 ns. The waiting time is characterized by small fluctuations of the end-to-end distance, during which small confor-



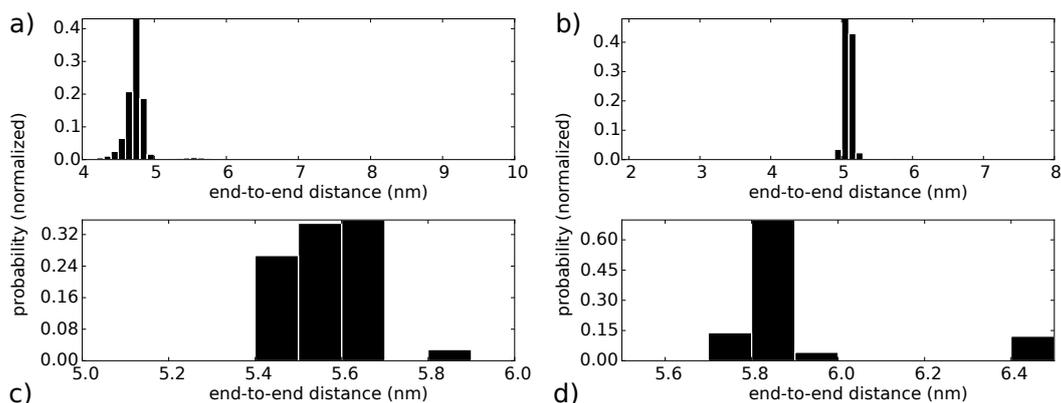
**Figure 4.2:** Typical evolution of the end-to-end distance during forced protein unfolding. In most cases, an initial rapid increase is followed by a plateau and a final rapid increase (red). For some samples, a plateau higher than the overall waiting time plateau is observed for part of the simulation (blue).

mational changes occur, but the protein remains native-like, with a mostly intact secondary structure. The waiting time usually represents the largest part of the overall unfolding time, and can vary between hundreds of picoseconds and hundreds of nanoseconds. In a few cases, this second phase cannot be accurately distinguished from the other two; the end-to-end distance increase is continuous, with only a middle region where the increase rate is reduced. The last phase starts from the waiting time characteristic plateau and ends with reaching the maximum length at which the simulations were stopped; like the initial increase, this phase is also short, taking in most cases below 2 ns. In a few cases, plateaus corresponding to end-to-end distances distinctly higher than the overall waiting time plateau are observed; the end-to-end distance increase between the overall waiting time plateau and the higher plateau, as well as between the higher plateau and the maximum length, is also short, below 1-2 ns, similar to the other increase phases. The evolution in time of the end-to-end distance in a set of 500 simulations for one value of the applied force is represented in Figure 4.3; the results are very similar for other forces, with the unfolding time scale showing a dependency on the force.

Due to the existence of flexible regions in both proteins, the end-to-end distance of the native structure is not unique. For ubiquitin, the most common values are in the range



**Figure 4.3:** Evolution of the end-to-end distance for a set of 500 simulations on ubiquitin at 664 pN (a) and NuG2 at 498 pN (b). Samples are ordered based on the unfolding time.



**Figure 4.4:** a,b) Histogram of end-to-end distances for a set of 500 simulations on ubiquitin (a) and NuG2 (b) unfolded at 664 pN and 498 pN, respectively. c,d) Histogram of end-to-end distances from all plateaus, weighted by the inverse of the simulation length, for ubiquitin (c) and NuG2 (d), unfolded at 664 pN and 498 pN, respectively. End-to-end distances below 5.0 nm, for ubiquitin (c), and 5.5 nm, for NuG2 (d), correspond to overall waiting time plateaus; higher distances correspond to unfolding semi-stable states.

3.5-3.8 nm; a higher variability is observed for NuG2, due to the longer initial flexible region. The overall waiting time plateau corresponds to lengths in the range 4.5-5 nm for ubiquitin and 4.5-5.5 nm for NuG2 (Figure 4.4a,b). The higher plateaus are more difficult to characterize due to their low probability of occurrence (Table 4.1). For ubiquitin, most such plateaus correspond to lengths in the range 5.0-6.0 nm (Figure 4.4c), with only a few in the range 6.0-7.0 nm observed at the lowest force (498 pN). For NuG2, these plateaus are only observed at the lowest forces and correspond to lengths in the range 5.5-6.5 nm (Figure 4.4d).

The higher end-to-end distance plateaus reflect a relative displacement of the  $\beta_1$  and  $\beta_5$  strands ( $\beta_1$  and  $\beta_4$  for NuG2) with respect to each other compared to the native structure. This displacement is accompanied by a decrease in the number and a rearrangement of the H-bonds between the two  $\beta$ -strands, as also shown in the characterization of an intermediate structure<sup>82</sup> or represented as yielding of residue contacts in a simplified model of ubiquitin<sup>83</sup>. The further displacement of these strands leads eventually to the loss of native contacts and complete unfolding. The order in which the various contacts between residues and secondary structure elements of ubiquitin break during the unfolding process has already been investigated<sup>82,83,91,92</sup>. We will focus on these aspects again in the later

Force (pN)	498	581	664	747	830	913	996
Ubiquitin (samples)	95	65	18	4	2	0	0
NuG2 (samples)	23	6	0	0	0		

**Table 4.1:** Frequency of occurrence of the higher end-to-end distance plateaus, indicating a semi-stable state. A total of 500 samples were unfolded for each force. For NuG2, the highest applied force was 830 pN.

part of this chapter, where the unfolding mechanism is investigated by means of analysis of internal forces in proteins.

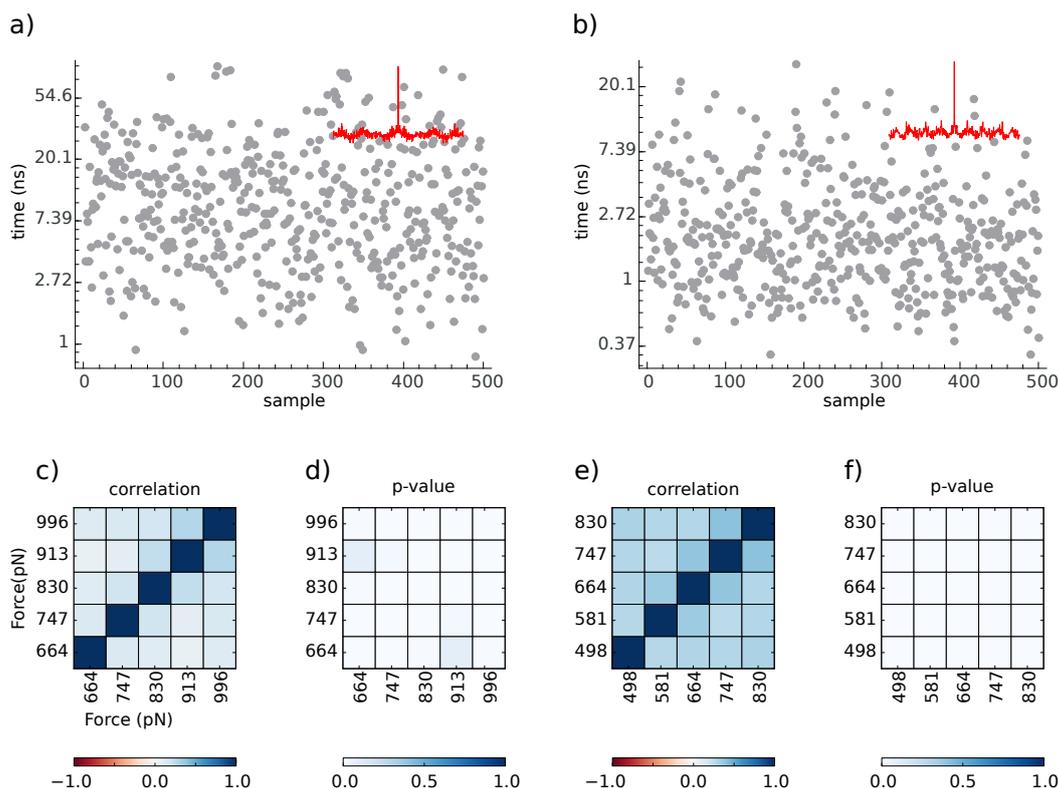
### 4.3 FORCE-CLAMP MD ERASES MOLECULAR MEMORY

The folding and unfolding of small or single-domain proteins are often considered Markovian processes, consisting of two states, memoryless and independent of the molecular history. It is however unclear if the unfolding determined by the application of an external force fulfills these conditions<sup>14</sup>. We investigate here whether the force-clamp MD simulations can lead to a loss of molecular history, and whether the path of transition between the two states depends on the applied force.

As mentioned in Section 2.1.2, an MD simulation starts with a set of atom coordinates and velocities, which are used to compute a new set of coordinates and velocities by integrating Newton's equations of motion. By repeating the integration, the molecular system evolves in time, and the different sets of atom coordinates which are thus obtained represent (possibly different) molecular conformations. The conformational space is huge, and only small parts of it are sampled during a typical MD simulation.

Regions of the conformational space characterized by a low potential energy form basins, surrounded by regions with higher energy, representing effective barriers. Thermal fluctuations of the atoms (Section 2.1.5) provide a small amount of kinetic energy, allowing the system to overcome barriers and sample the low energy basins. Large barriers cause the system to become trapped in a very limited region of the conformational space. Thus, the starting coordinates and velocities have a strong influence on the conformations which can be visited during the simulation. The starting coordinates are often obtained by saving conformations at regular intervals from a long equilibrium simulation; the longer the simulation, the higher the probability of generating starting coordinates which could theoretically correspond to different basins on the energy landscape. The starting velocities are randomly generated from a Boltzmann distribution corresponding to the desired temperature; their randomness helps towards a better sampling of the conformational space, even when the initial coordinates are not different enough.

The application of an external force can induce a molecular system transition between basins, such as the basins of the folded and unfolded states, by lowering the energy barrier (Section 1.2). But can the external force also influence the way the transition happens? We attempt to answer this question by using the large amount of MD simulations performed at a single value of the force, as well as simulations started from the same initial state on which we applied different forces. As a measure of similarity between the simulations, we use the unfolding time. Running several MD simulations starting from the same set of



**Figure 4.5:** Unfolding times obtained from a set of 500 samples subjected to an external force of 664 pN (a) or 830 pN (b). The autocorrelation function (red inset) shows that unfolding times are independent. Spearman rank-order correlation coefficients (c, e) and associated p-values (d, f) for the same 500 samples of ubiquitin (c, d) and NuG2 (e, f) subjected to different external forces.

initial coordinates and velocities leads to the same evolution of the molecular system and therefore to the same unfolding time. A difference in the unfolding times thus indicates one or more of the following: a different set of initial coordinates, a different set of initial velocities, or an influence of the external force on the evolution of the system. It is however not possible to distinguish between these different causes from the analysis of the unfolding times alone.

We observe no correlation between the unfolding times of successive samples from a set of 500 MD simulations of ubiquitin subjected to an external force of 664 pN (Figure 4.5a). The same randomness of unfolding times can be observed at all forces applied here (for example at 830 pN, Figure 4.5b) or for other proteins (NuG2, data not shown). This indicates that the simulation protocol used to obtain these results is able to erase the molecular memory, but does not single out which of the steps of the protocol is responsible. (The simulation protocol is detailed in Appendix C and starts with the Protein Data Bank structures 1UBQ and 1MI0, for ubiquitin and NuG2, respectively).

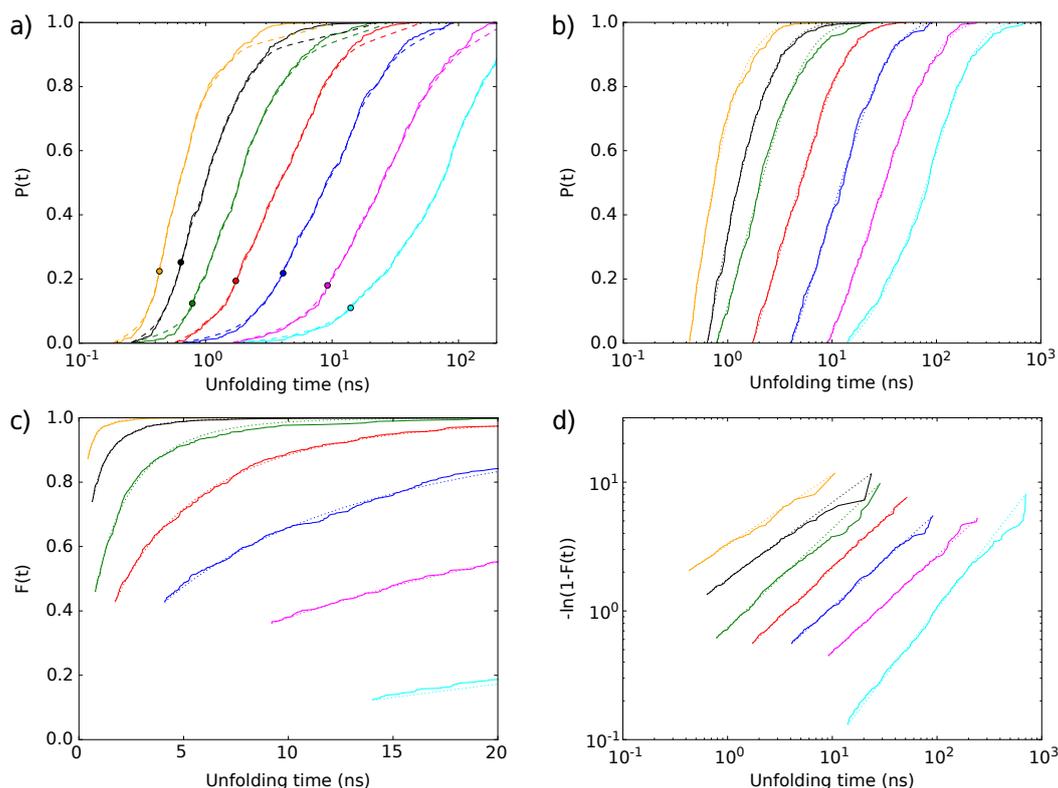
For a single sample, characterized by a set of initial coordinates and velocities, we can apply external forces of different magnitudes to study their influence on the evolution of

the molecular system. As the initial state of the system is identical, any variability in the unfolding times is necessarily due to the variability in the value of the force. However, a direct comparison of unfolding times for a single sample is not relevant; it is expected that a higher force will result in a shorter unfolding time due to the lowering of the energy barrier (Section 1.2). Instead, we compare the unfolding time in relation to the unfolding times of other samples. If the external force does not influence the evolution of the system, a sample which unfolds faster or slower compared to other samples is expected to do so at all forces. From Figures 4.5a and b, we can already observe that the unfolding times of the same sample at different forces are not correlated; samples which unfold fast at 664 pN might unfold slow at 830 pN and *vice versa*. This result is not limited to the two forces; the low correlation coefficients (Figures 4.5c and d) extend it to the whole range of forces used here. Furthermore, we obtain a similar picture for another protein (NuG2, Figures 4.5e and f), suggesting that this is a universal behavior in the mechanical unfolding of proteins. We can therefore answer the above question - the application of the external force can indeed change the evolution of the molecular system, erasing its memory.

#### 4.4 NON-EXPONENTIAL UNFOLDING KINETICS

The Arrhenius equation has long been used to describe the kinetics of simple two-state reactions. It relies on the assumption that the reaction has a single transition state and a well defined barrier between states, determining the reaction rate. The unfolding of small or single-domain proteins has been assumed to be such a two-state process, with the survival probability following a single exponential time dependency. This assumption further implies that there is a single unfolding pathway, and that the protein has a single native conformation or, more generally, that the conformations interconvert much faster than the unfolding time. However, experimental evidence on ubiquitin and the I27 domain of titin<sup>3,4</sup> has revealed a non-exponential survival probability, which has been best described by a stretched exponential model<sup>3,84</sup>. As a consequence, it has been proposed that the single native conformation should be replaced by an ensemble of conformations which interconvert on time scales similar to, or larger than, the unfolding time, and that the barrier crossing reaction is similar to glassy dynamics<sup>3</sup>.

Later, it was found that a single exponential best models experimental data from forced unfolding of NuG2<sup>14</sup>. This was a surprising result, given the structural resemblance of ubiquitin and NuG2. Only recently, a detailed analysis of the differences between the two sets of experiments has revealed that the cause of the contrasting outcome is an alternative way of interpreting the results<sup>84</sup>.



**Figure 4.6:** Kinetic profiles from forced unfolding of ubiquitin. Dotted lines represent stretched exponential fits. a) Empirical CDF has an S-shaped profile for all forces, the inflection point (represented as a circle) defines  $t_{\min}$ . b) Unfolding profiles after elimination of points below  $t_{\min}$ . c) Modified CDF after elimination of points below  $t_{\min}$ . d) Same as c), with vertical axis representing  $-\ln(1-F(t))$ , resulting in a straight line for a stretched exponential.

From force-clamp MD simulations, we obtain unfolding times for sets of 500 samples subjected to the same force. To avoid inconsistencies, we analyze our data following the procedure recommended for the interpretation of the experimental data<sup>84</sup>. We start by building an empirical cumulative distribution function (CDF),  $P(t)$ , representing the unfolding probability as a function of time. By sorting the unfolding times and ranking them from the smallest to largest, we obtain an empirical CDF, which expresses the variation of the unfolding time as a function of the rank in the set:

$$P(t) = n_t/N \quad (4.1)$$

where  $n_t$  is the number of samples unfolded before time  $t$ , and  $N$  is the total number of samples in a set (500). This empirical CDF has an S-shape for all forces used in this study (Figure 4.6a), and for both ubiquitin and NuG2.

The dynamics of the unfolding process can offer an explanation for the observed S-shaped profile. It cannot be expected that one or more samples unfold in virtually zero time; this would be equivalent to some protein atoms moving over distances of several

nanometers instantaneously, which is unrealistic. Instead, after applying the external force on the C- and N-terminal residues, the protein elongates with a finite, variable speed; this is expressed through an increase in the end-to-end distance. There are two processes which can be seen as opposing this elongation: an internal friction between parts of the protein which rearrange, and an external friction between protein and water, as water needs to be displaced to make space for the elongated protein. These processes can slow down the elongation, such that the probability of an unfolding event happening during this time is reduced or, equivalently, fewer than expected unfolding events can happen. The initial part of the S-shaped curve, showing a slowly increasing unfolding probability, captures this effect. Theoretically, if this effect would be missing, the initial part of the S-shape curve would disappear; this could be achieved by using an umbrella potential (Section 2.1.6) to slowly pre-elongate the protein, allowing the external force to distribute throughout its structure. However, the conformational changes associated to this process could already start the unfolding of the protein, leading to a non-realistically high unfolding probability. To eliminate the influence of this effect on the overall kinetic curve, we determine a minimum time  $t_{\min}$  below which samples show the slow increasing probability, and remove those samples which complete their unfolding faster than  $t_{\min}$ . The excluded samples are valid unfolding events but, as mentioned above, their number is lower than expected during this time interval. The slowly increasing probability allows us to identify  $t_{\min}$  as the inflection point of the kinetic curve. To determine it, we initially apply a smoothing to the kinetic curve, followed by numerical differentiation and finding of the minimum (Figure 4.6a and Table 4.2).

We also note that a similar procedure needs to be applied for the interpretation of the experimental data, where events at both ends of the kinetic curve have to be discarded. At the lower end, a parameter similar to  $t_{\min}$  is determined by the frequency of the AFM constant force feedback mechanism, such that only unfolding times higher than a few tens of milliseconds can be recorded. At the higher end of the curve, dissociation of the protein from the indenter tip or from the substrate can terminate the experimental observations sooner than expected, and too few unfolding events are detected. As a consequence, the analysis of the unfolding traces for multi-domain protein constructs used in most experiments is quite complex and prone to interpretation bias<sup>84</sup>.

Interestingly, a similar S-shaped kinetic curve has been previously observed for the mechanical unfolding of ubiquitin<sup>93</sup>. Fitting the kinetic curves obtained from Langevin dynamics simulations involving a simplified protein model required a double exponential for forces even lower than those studied here. Yew *et al.*<sup>93</sup> also proposed that the free-energy profile might have a more complex shape than those assumed by the Bell or DHS models

(Section 1.2), or might even be multidimensional. Their theory agrees well with that of static disorder<sup>5</sup>, and their findings qualitatively agree with our results.

After discarding the samples which unfold below  $t_{\min}$ , the empirical CDF has the expected exponential-like shape (Figure 4.6b). A stretched exponential (also known as Weibull) model has been shown to best fit the kinetic data obtained from experiments<sup>84</sup> and will also be used here. It can be expressed as:

$$F(t) = 1 - e^{-(\alpha t)^b} \quad (4.2)$$

where  $F(t)$  is the theoretical or true CDF of unfolding times,  $t$  is the unfolding time of a single sample,  $\alpha$  is the unfolding rate constant, and  $b$  is an exponential constant. We note that the stretched exponential model also covers the single exponential case, for which  $b$  is 1.

If we consider  $t_{\max}$  to be the largest recorded unfolding time for each force, the empirical CDF can be expressed as:

$$P(t) = F(t)/F(t_{\max}) \quad (4.3)$$

However, because samples before  $t_{\min}$  were discarded, their contribution should be removed, such that  $F(t)$  is shifted by  $F(t_{\min})$ :

$$P(t) = \frac{F(t) - F(t_{\min})}{F(t_{\max}) - F(t_{\min})} \quad (4.4)$$

where  $t$  is always bracketed by  $t_{\min}$  and  $t_{\max}$ . If the regime of slow unfolding probability below  $t_{\min}$  would not exist (in other words,  $t_{\min} = 0$ ), the two expressions of  $P(t)$  above are the same.

The results of fitting Equation 4.4 to our data can be seen in Figure 4.6b; the corresponding parameter values are listed in Table 4.2. The fits are very good for the whole range of times between  $t_{\min}$  and  $t_{\max}$ , and for all forces. Having the parameters determined from the fit allows us to evaluate  $F(t_{\min})$  and  $F(t_{\max})$ , and to rewrite the theoretical CDF to remove the contribution of the samples unfolding faster than  $t_{\min}$  as:

$$F(t) = P(t)(F(t_{\max}) - F(t_{\min})) + F(t_{\min}) \quad (4.5)$$

with the results represented in Figure 4.6c. To better evaluate the goodness of fit, Figure 4.6d uses a different scale, in which a stretched exponential would be represented as a straight

Force (pN)	a (ns <sup>-1</sup> )	b	Inflection point (ns)	% of data after inflection point
498	0.010	1.00	14.02	89.2
581	0.037	0.75	9.20	82.3
664	0.110	0.74	4.10	78.4
747	0.271	0.77	1.73	80.8
830	0.668	0.77	0.79	87.8
913	2.577	0.60	0.64	75.0
996	8.805	0.54	0.43	77.8

**Table 4.2:** Parameters obtained from fitting the stretched exponential model to the empirical CDF of unfolding times for ubiquitin at various forces.

Force (pN)	a (ns <sup>-1</sup> )	b	Inflection point (ns)	% of data after inflection point
498	0.049	0.80	4.34	85.2
581	0.190	0.72	1.84	85.2
664	0.460	0.75	0.81	87.0
747	1.311	0.66	0.53	83.8
830	1.726	0.99	0.29	90.0

**Table 4.3:** Parameters obtained from fitting the stretched exponential model to the empirical CDF of unfolding times for NuG2 at various forces.

line with the slope equal to the exponential constant  $b$  (Equation 4.2). Again, if the regime of slow unfolding probability below  $t_{\min}$  would not exist, Equation 4.5 would become:

$$F(t) = P(t)F(t_{\max}) \quad (4.6)$$

As  $P(t)$  starts at zero (or very close to it),  $F(t)$  should also start at zero; however, Figures 4.6c and d show  $F(t)$  to start at values ranging between 0.15 and 0.9 for the different forces. The missing curve segments between zero and the observed starting values are due to the samples with unfolding times below  $t_{\min}$ , not captured by our kinetic curves. It is interesting to note the amount of unfolding events to which they correspond: Table 4.2 shows that the number of samples discarded from the empirical CDF varies between 10 and 25%, while Figure 4.6c estimates that 15–90% of the theoretical unfolding events predicted by a stretched exponential model would happen below  $t_{\min}$ .

The values of the exponential constant  $b$  show a decreasing trend as the forces become higher (Tables 4.2 and 4.3), for both ubiquitin and NuG2. The same can be observed by applying the fitting procedure described above for kinetic data obtained from unfolding under force of single molecules of ubiquitin<sup>4</sup> (Lannon and Brujić, personal communication).

This is equivalent to a flattening of the stretched exponential curve with increasing force, suggesting the two-state kinetic model of protein unfolding should be augmented by a component expressing the elastic behavior of the protein stretched by an external force. The single exponential model, assumed so far for protein unfolding under any conditions, would thus remain valid only for spontaneous unfolding in the absence of force. Yew *et al.*<sup>93</sup> reached a similar conclusion, although using a different model. Unfortunately, the data that we have so far does not allow us to derive such an elastic dependency of the unfolding kinetics. Apart from having only a few values of the  $b$  exponential constant, they are also strongly linked to the exact shape of the empirical CDF, which is influenced by the choice of  $t_{\min}$ . More samples from which to determine the empirical CDF as well as more values of the external force would be required before such a dependency could be derived.

The hypothesis that the unfolding kinetics under force depends on the elasticity of the protein is further supported by the correlation between the exponential constant  $b$  and the mechanical resistance of the protein. Previous studies of force spectroscopy<sup>4,14</sup> have found that the I27 domain of titin is more mechanically stable than ubiquitin, which in turn is more stable than NuG2. For the lower forces used in our study, the values of the  $b$  exponential constant are systematically larger for ubiquitin than for NuG2. New experimental data from forced unfolding of I27 and ubiquitin shows higher  $b$  exponential constant values for I27 than for ubiquitin (Lannon and Brujić, personal communication), rounding up our results. We therefore suggest that the unfolding kinetic curve becomes more stretched (or flattened) with the decreasing mechanical resistance of the protein. According to this hypothesis, a mechanically stable protein or domain, like I27, would transition more suddenly between the native and unfolded states, closer to a two-state kinetic model, such that the unfolding probability profile would resemble a single exponential. On the contrary, a protein with low mechanical resistance, like NuG2, would show a rubber-like behavior, with the transition between native and unfolded states delayed by the protein elasticity.

The structural analysis (Section 4.2) has revealed the existence of plateaus in the end-to-end distance evolution which are higher than the overall waiting time plateau. These plateaus correspond to semi-stable states; their existence time scale ranges from hundreds of picoseconds to tens of nanoseconds. However, there was no case in which the sample transitioned back from such a semi-stable state to the overall waiting time plateau, indicating that these are intermediary structures along the unfolding pathway. But do these semi-stable states influence the unfolding kinetics? More precisely, is there a correlation between the existence of the semi-stable states and the stretched exponential kinetics?

If the semi-stable state would be present in samples with high unfolding times with higher probability than in samples with low unfolding times, it could indeed explain the

stretched exponential shape of the kinetic profile. The semi-stable state would act in fact as an additional state in the transition between the folded and the unfolded states, thereby slowing down the process. However, the semi-stable states appear randomly distributed among the entire set of samples; there is no correlation between their presence or length and the unfolding time of the samples.

The fact that the semi-stable states are not directly related to the stretched exponential kinetics is further supported by the stretched exponential kinetics observed at higher forces, where the semi-stable states are absent. We performed an additional test for the set subjected to a 664 pN force by looking for a stretched exponential unfolding profile after splitting the samples in two sets: one for the samples containing the semi-stable state and one for the samples without it. As there were only 18 samples showing semi-stable states, the data set was too small for reliable fitting, and was extended to 39 samples by relaxing the criterion such that all samples showing an end-to-end distance between 5 and 6 nm for at least 1 ns were made part of this set. Both sets could be well fitted by a stretched exponential, showing again that the presence of the semi-stable state is not correlated to a change in kinetic behavior.

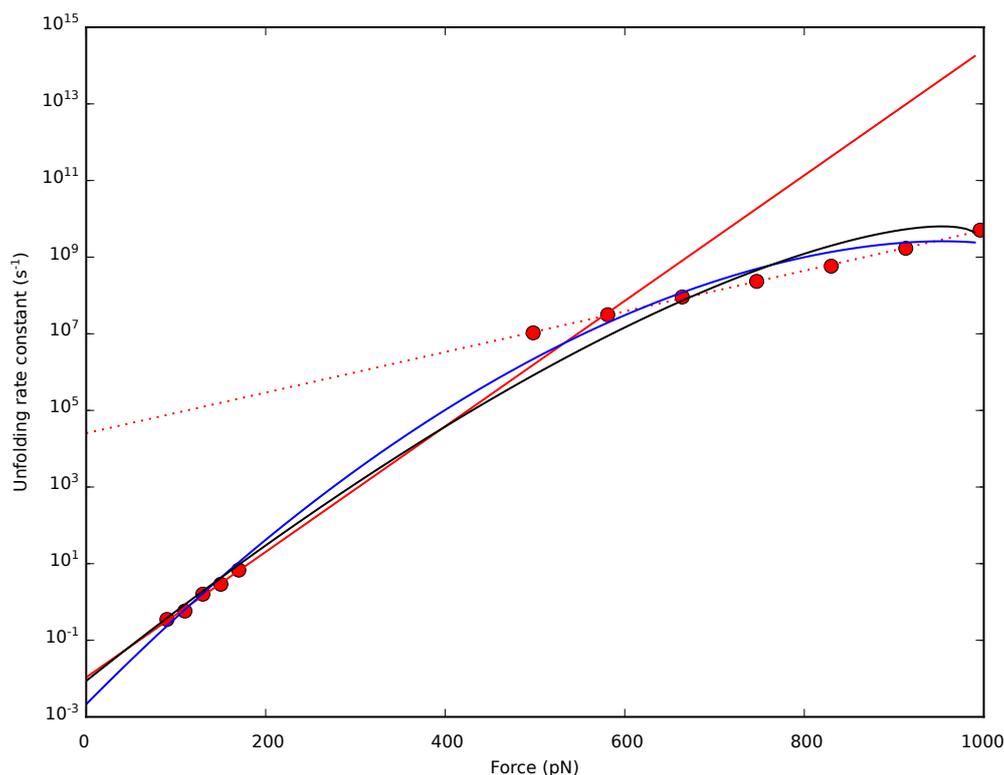
Higher plateaus in the evolution of the end-to-end distance can also be observed in unfolding simulations of NuG2 for the lower forces (498 and 581 pN). As in the case of ubiquitin, the semi-stable state is randomly distributed among the samples, thus its presence cannot explain the stretched exponential kinetics. We therefore conclude that the semi-stable states can only be considered a feature of the static disorder in unfolding pathways of ubiquitin under force<sup>5</sup>.

#### 4.5 KINETIC PARAMETERS IN THE ABSENCE OF FORCE

Having determined the unfolding kinetic parameters  $a$  and  $b$  for the different forces, we can use them to extrapolate the unfolding rate constant in equilibrium (at zero force,) and to estimate the distance along the unfolding reaction coordinate between the reactant, native-like state, and the transition state. The Bell and DHS models predict that the unfolding rate constant depends on the force acting on the protein (Section 1.2); by lowering the free energy barrier, the application of force induces a higher rate of unfolding. In addition to the unfolding rate constant and the distance to the transition state, the DHS model also estimates the free-energy barrier.

The characteristic unfolding rate depends on the model describing the kinetics. For the stretched exponential model used here, it is defined as:

$$\alpha = \frac{1}{\int_0^{\infty} 1 - F(t) dt} = \frac{1}{\int_0^{\infty} e^{-(at)^b} dt} \quad (4.7)$$



**Figure 4.7:** Determining equilibrium kinetic parameters for ubiquitin from unfolding rate variation as a function of force. Bell model fit for experimental data (solid red); Bell model fit for MD data (dotted red); DHS model fit for combined data, assuming a cubic (black) and cusp (blue) energy profiles. Unfolding rates were derived from experimental data (Lannon and Brujić, personal communication) for forces below 200 pN, and from MD data for forces starting at 498 pN.

which has the form:

$$\alpha(a, b) = \frac{a}{\Gamma(1 + 1/b)} \quad (4.8)$$

where  $\Gamma(x)$  is the gamma function. Figure 4.7 shows Bell model fits for  $\alpha$  values calculated from our data and, separately, from experimental data<sup>4</sup> (Lannon and Brujić, unpublished results). In addition, we represent a DHS model fit to the experimental and our data combined. The parameters obtained from these models are listed in Table 4.4.

As mentioned in Section 1.2, the Bell model has been widely used in previous studies of protein unfolding under constant force. Its validity should hold for similar ranges of forces, and Figure 4.7 shows indeed that it models the data well when MD or experimental data are used separately for fitting. However, it is clear that it cannot fit well a combined set containing both the MD and the experimental data. This is also evident from the comparison of zero force unfolding rate constants ( $\alpha_0$ ) and distances to transition state ( $\Delta x$ ). While  $\Delta x$  obtained from MD data is 3 times smaller than the one obtained from experimental data, the unfolding rate constants have widely different orders of magnitude. The

	$\alpha_0$ ( $s^{-1}$ )	$\Delta x$ (nm)	$\Delta G$ (pNnm)	Critical force (pN)
Bell model, MD data	$2.5 \cdot 10^4$	0.03		
Bell model, experimental data	0.011	0.09		
Cubic DHS model ( $\nu = 2/3$ )	0.009	0.11	73	1004
Cusp DHS model ( $\nu = 1/2$ )	0.002	0.14	76	1095

**Table 4.4:** Kinetic parameters obtained from fitting the Bell and DHS models to MD and experimental data for ubiquitin.

	$\alpha_0$ ( $s^{-1}$ )	$\Delta x$ (nm)	$\Delta G$ (pNnm)	Critical force (pN)
Bell model, MD data	$2 \cdot 10^4$	0.03		
Bell model, experimental data	0.023	0.16		
Cubic DHS model ( $\nu = 2/3$ )	0.135	0.11	64	837
Cusp DHS model ( $\nu = 1/2$ )	0.059	0.14	66	919

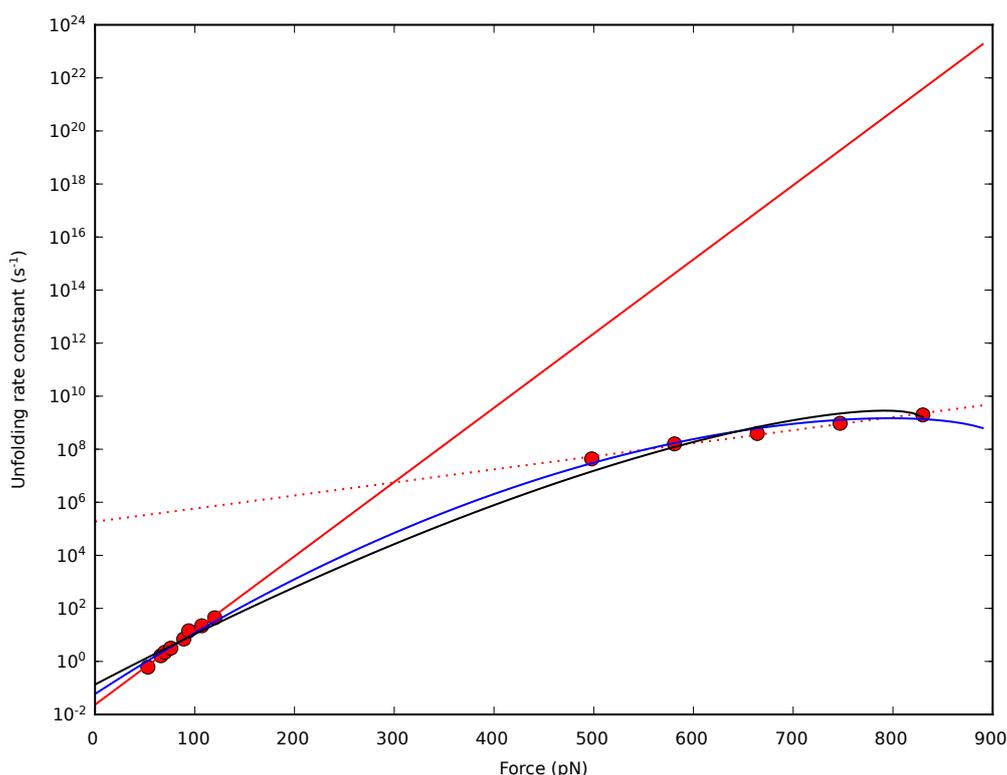
**Table 4.5:** Kinetic parameters obtained from fitting the Bell and DHS models to MD and experimental data for NuG2.

$\alpha_0$  value obtained from the MD data is unrealistic, indicating a high rate of spontaneous unfolding events which is not observed in natural conditions. We note that the  $\alpha_0$  value obtained from the experimental data is different from the one obtained in experiments of chemically induced denaturation<sup>94</sup> ( $4.3 \cdot 10^{-4} s^{-1}$ ), however it was already proposed<sup>13</sup> that this difference appears due to different reaction mechanisms in the two types of experiments. As the experimental data was generated for a small range of small forces, we can assume the Bell model to hold and the  $\Delta x$  and  $\alpha_0$  values obtained from it to be realistic.

One of the goals in developing the DHS model is bridging the gap in forces and time scales between experiments and MD simulations of mechanical unfolding of proteins. Applying it to the combined MD and experimental data and using the parameter  $\nu = 2/3$  corresponding to a cubic energy landscape, we obtain  $\Delta x$  and  $\alpha_0$  values very similar to the ones predicted by the Bell model from the experimental data. Choosing the other recommended value of  $\nu = 1/2$  corresponding to a cusp-like barrier in the energy landscape, the fitting curve does not look significantly different, but the resulting unfolding rate constant is several times smaller and the distance to the transition state is around 25% larger compared to the cubic model. The predicted free-energy barrier is comparable between the two models, differing by around  $k_B T$ , and in both cases the maximum forces used are lower than the critical forces defined for each model. We conclude that the cubic model of the energy landscape is the most appropriate for ubiquitin. We note, however, that the individual Bell model fits to the experimental and MD data perform better than the DHS fit. Each of the two data sets follows an approximately linear dependency on the force on the semi-log

plot, while a better DHS fit could only be obtained if this dependency would show a relative curvature. Thus, we cannot exclude the possibility that our simulations probe an unfolding mechanism and barrier different from the experiments, a question which can ultimately be answered only by closing the gap in forces and time scales between experiments and MD simulations.

The same analysis can also be applied to MD data and experimental results<sup>14</sup> describing the mechanical unfolding of NuG2 (Figure 4.8, Table 4.5). Similar to ubiquitin, the Bell model is able to fit well the MD and experimental data separately, but not the combined set. The equilibrium  $\Delta x$  and  $\alpha_0$  derived from MD and experimental data are also significantly different, with the unfolding rate constant from MD simulations predicting an unrealistically high rate of spontaneous unfolding events. The DHS model applied on the combined set of data requires  $\nu = 1/2$  in order to reach values of the kinetic parameters comparable to those obtained from the Bell model applied to experimental data; the similarity is however lower than for ubiquitin. By comparison, the DHS model using  $\nu = 2/3$  predicts a too high  $\alpha_0$  and a too low  $\Delta x$ , even though the estimated free-energy barriers differ by less



**Figure 4.8:** Determining equilibrium kinetic parameters for NuG2 from unfolding rate variation as a function of force. Bell model fit for experimental data (solid red); Bell model fit for MD data (dotted red); DHS model fit for combined data, assuming a cubic (black) and cusp (blue) energy profiles. Unfolding rates were derived from experimental data<sup>14</sup> for forces below 200 pN, and from MD data for forces starting at 498 pN.

than  $k_B T$ . In conclusion, the cusp model of the energy landscape seems the appropriate one for NuG2, in contrast to ubiquitin.

#### 4.6 INTERNAL DISTRIBUTION OF FORCES AND STRESSES

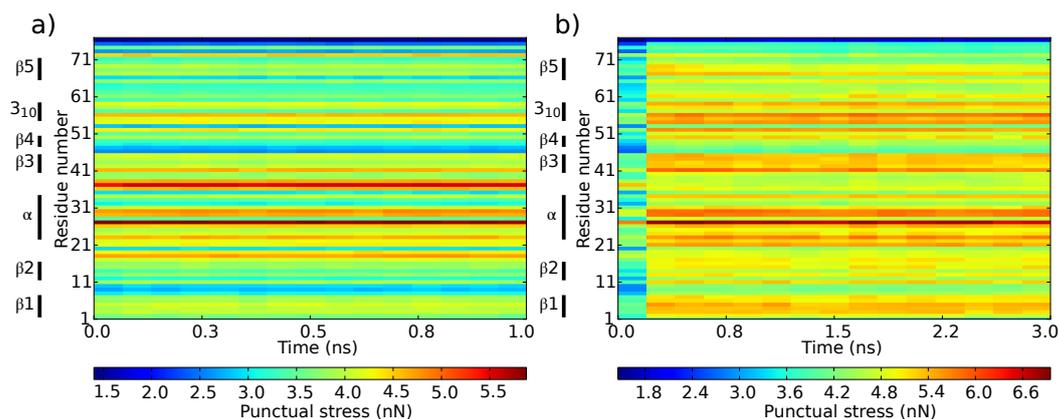
Even in the absence of an external force, a protein is not a relaxed structure. The structural elements of the protein can be found in tensed or compressed states which balance each other, giving rise to a pre-stress<sup>41</sup>. Internal forces between different parts, often distant, of the molecule form a mesh that maintains the overall shape. Such forces are large enough to prevent the structure of the protein from being changed by the thermal fluctuations of its own atoms or of atoms of the solvent. It is also conceivable that these forces oppose the deformation of the protein under the effect of external perturbations, thus possibly influencing the unfolding process. In this section, we investigate the link between the internal stress and force distribution and the stretched exponential kinetics describing the mechanical unfolding of ubiquitin and NuG2.

##### 4.6.1 *Distribution and propagation of mechanical stress*

The punctual stress expresses the accumulation of forces on an atom or residue (Section 2.3.3). It is therefore an ideal measure for the internal interactions of a protein, highlighting the structural elements which cooperate in maintaining the geometry of the molecule. Below, we use TRFDA to compute the punctual stress, investigating how the internal balance of forces changes as a result of applying an external mechanical load.

To make such a comparison, we start by investigating the stress at residue level in the absence of an applied force. As the punctual stress can show large fluctuations (Section 2.2), we reduce the noise by averaging the stress over 500 equilibrium simulations used to produce the starting structures for the forced unfolding (Figure 4.9a). The range of stress values is large; the highest value is more than three times larger than the lowest one, and reaches magnitudes of around 20% of the rupture stress in a very strong material (Figure 3.11), in stark contrast with the classical view of a relaxed protein.

We observe that the stress on each residue remains almost constant during the entire equilibrium simulations. This suggests that the interplay of internal forces remains unaltered, conferring mechanical stability to the protein. Low stress residues are mostly located within the flexible regions of the molecule, where they do not participate in maintaining a certain molecular geometry. In contrast, residues in or around the helices show high levels of stress, suggesting that these are important structural elements. A surprisingly high stress is found in the middle of the loop between the  $\alpha$ -helix and  $\beta_3$  strand; this occurs



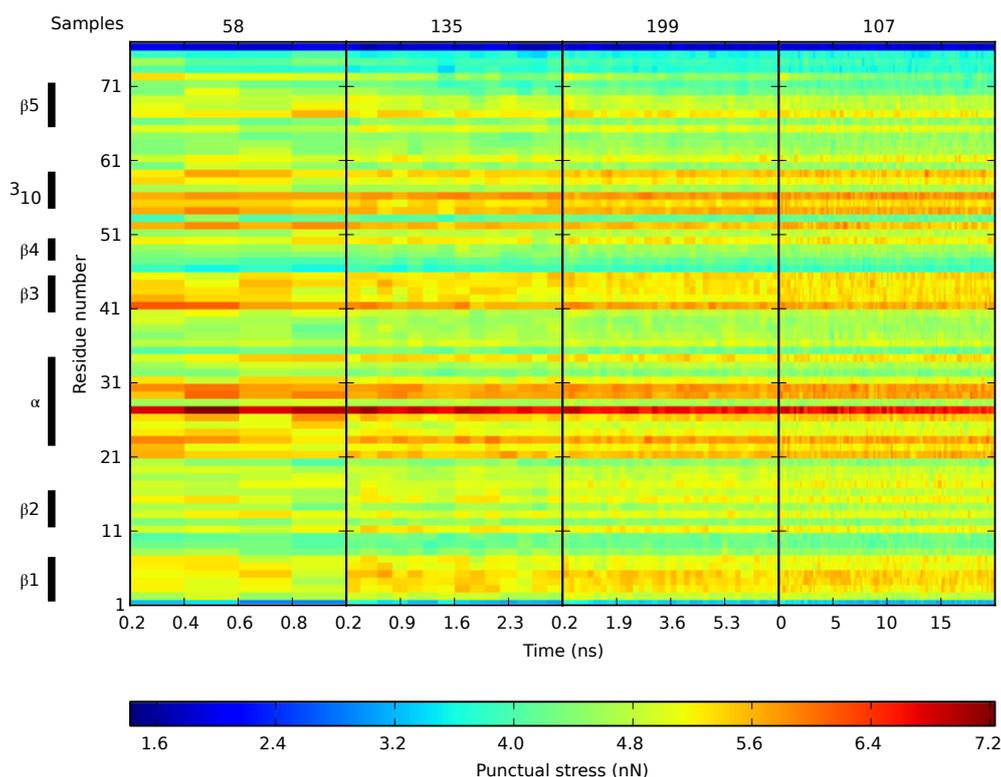
**Figure 4.9:** Average residue punctual stress during equilibrium simulations (a) and during the first 3 ns after applying an external force of 664 pN.

due to a salt bridge formed with the C-terminal residue<sup>41</sup>, suggesting that the flexible tail of the protein remains in equilibrium close to the core of the protein.

Upon application of a constant force on the two terminal residues, the stress levels increase, however the distribution is different from equilibrium (Figure 4.9b). High levels of stress are more widespread, in particular on most of the  $\beta$ -strands. Although the force application points are directly attached to, or part of, the  $\beta 1$  and  $\beta 5$  strands, the stress on these strands increases only moderately. In contrast, the stress on the  $\beta 3$  strand and the  $3_{10}$ -helix increases significantly, indicating that the mesh of internal interactions is very effective in dissipating the external force. This is analogous to the dissipation taking place in a bridge or the wing of an airplane, structures which thereby are able to withstand large forces, suggesting that the mechanical resistance was subject to optimization, possibly through an evolutionary mechanism<sup>95</sup>.

Figure 4.9b also shows that the stress levels remain almost constant during the waiting time, indicating again a stable interplay of the internal interactions which oppose the deformation. In contrast, the initial evolution from pre-stress to these constant levels is very fast, taking place in less than 200 ps. This agrees well with a previous study which found force to propagate on a picosecond timescale in a poly-Ala chain<sup>96</sup>. This is a surprising result, as the protein unfolding in our force-clamp simulations occurs nanoseconds to hundreds of nanoseconds after the application of the force, showing a discrepancy of 3-5 orders of magnitude. This suggests that the external force mostly weakens the interactions stabilizing the protein structure, while the main factor determining the beginning of the unfolding process is stochastic - the thermal fluctuations of the atoms.

To investigate this hypothesis, we split the 500 samples into groups, based on the length of their waiting times, and then average the stress in each group (Figure 4.10). We observe that the stress levels are very similar between the groups, suggesting that the distribution



**Figure 4.10:** Average residue punctual stress during the waiting time, after applying an external force of 664 pN. Samples were grouped based on their waiting time. First group (58 samples) has waiting times between 1 and 3 ns. Second group (135 samples) has waiting times between 3 and 7 ns. Third group (199 samples) has waiting times between 7 and 20 ns. Fourth group (107 samples) has waiting times larger than 20 ns.

of the external force throughout the structural elements of the protein is approximately the same for both the samples that unfold fast and for those with a long waiting time. Thus, the stress distribution cannot provide an explanation for the stretched unfolding profiles we observed (Section 4.4). However, the similar levels of stress in both fast and slow unfolding samples are not an indication of a unique unfolding path, and therefore of a single exponential kinetics. In particular, there is no evidence of a particular stress distribution that is required before the start of the unfolding process; this supports the idea of glassy dynamics during forced unfolding<sup>3</sup>, which predicts the existence of an ensemble of unfolding paths rather than a single path. We note though that here the stress is recorded at time intervals of 200 ps. It is therefore possible that our data misses a fast redistribution of stress, similar to the one that happens at the onset of the external force, and which occurs again at the very moment the unfolding process begins. This hypothesis is subject of future studies. We note that similar results are obtained from equilibrium and force-clamp simulations of NuG2, respectively.

The punctual stress is very useful for quickly highlighting the “hot spots” of the stress distribution in the molecules. However, it is defined as a sum of pairwise forces, and thus

looses the details of the individual interactions. Below, we turn to using pairwise forces directly, for studying hydrogen bonds and the distribution of internal forces during the early stages of the unfolding process.

#### 4.6.2 Forces in hydrogen bonds

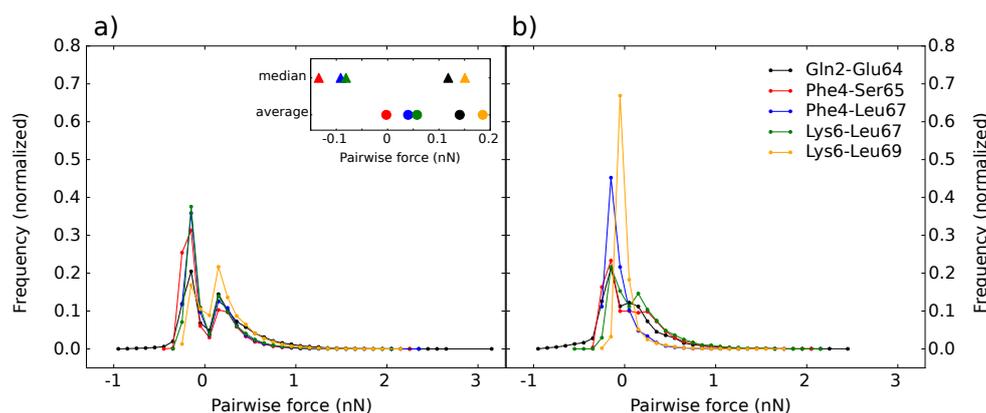
There are five backbone hydrogen bonds between the  $\beta_1$  and  $\beta_5$  strands of ubiquitin<sup>91</sup>. Table 4.6 lists their probability of being formed during our equilibrium simulations. Most H-bonds exist for well over 90% of the time; the Lys6-Leu69 H-bond, located towards the C-terminal end of the  $\beta_1$ - $\beta_5$  interface, exists for about 2/3 of the time - still a significant fraction. This indicates a stable arrangement, not only of the two strands, but also of the surrounding protein structure.

Residues	Probability (%)
Gln2(CO)-Glu64(NH)	93.0
Phe4(NH)-Ser65(CO)	96.8
Phe4(CO)-Leu67(NH)	98.3
Lys6(NH)-Leu67(CO)	98.2
Lys6(CO)-Leu69(NH)	67.3

**Table 4.6:** Probability of existence of H-bonds between the  $\beta_1$  and  $\beta_5$  strands of ubiquitin during equilibrium simulations.

Although the probabilities of existence of H-bond can offer insights into the chain of events surrounding the mechanical unfolding of ubiquitin<sup>91,92</sup>, the analysis of forces between the residues forming these H-bonds can more directly indicate their involvement in the unfolding process. Such analysis also allows us to determine whether the H-bonds between the  $\beta_1$  and  $\beta_5$  strands alone are responsible for the mechanical resistance of the protein, or if other structural elements are involved as well.

Figure 4.11a shows the fluctuations of pairwise forces between residues forming H-bonds in equilibrium simulations of ubiquitin. All H-bonds appear to sample both positive (repulsive) and negative (attractive) values with similar probabilities, and with significantly larger values on the positive side. There is a significant decrease in frequency around the zero force for all H-bonds, suggesting that they are found in either a compressed or tensed state for most of the time. The H-bond Lys6-Leu69 has a lower lifetime and has been found to break earlier than the others during mechanical unfolding<sup>91</sup>; this can be explained by its larger frequency of positive (repulsive) values compared to the negative ones, and to the low negative values it samples, suggesting that the bond is mostly in a compressed state and does not resist much to tension. Overall, the distribution of the forces along the  $\beta$ -strands shows all H-bonds reaching highly compressed states with similar probabilities, however the attractive regime seems to increase towards the N-terminal end.



**Figure 4.11:** a) Fluctuations of pairwise forces between H-bonds forming residues located on the  $\beta_1$  and  $\beta_5$  strands of ubiquitin during equilibrium simulations. Inset: comparison between the average (squares) and median (circles) values for the pairwise forces in a). b) Fluctuations of pairwise forces between H-bonds forming residues located on the  $\beta_1$  and  $\beta_5$  strands of ubiquitin during force-clamp MD simulations at 830 pN.

Residues	Probability (%)
Lys <sub>10</sub> (CO)-Phe <sub>58</sub> (NH)	91.3
Val <sub>12</sub> (NH)-Phe <sub>58</sub> (CO)	94.0
Val <sub>12</sub> (CO)-Thr <sub>59</sub> (OH)	34.9
Val <sub>12</sub> (CO)-Val <sub>60</sub> (NH)	40.9
Val <sub>14</sub> (NH)-Val <sub>60</sub> (CO)	50.2
Val <sub>14</sub> (CO)-Glu <sub>62</sub> (NH)	13.7

**Table 4.7:** Probability of existence of H-bonds between the  $\beta_1$  and  $\beta_4$  strands of NuG2 during equilibrium simulations.

For some of the H-bonds, the two measures are similar, but, for others, one measure indicates a compressed state, while the other a tensed state. In an earlier study on pre-stress carried out on ubiquitin by computing average pairwise forces<sup>41</sup>, all H-bonds between the  $\beta_1$  and  $\beta_5$  strands were found in a compressed state. Our results expose a rather different picture, suggesting that a more complete statistical analysis of the pairwise forces from equilibrium simulations should be performed before describing them with a single parameter, such as the average value. This is, however, beyond the scope of this work, and will be the subject of a future study.

Nevertheless, in our force-clamp MD simulations, we can continue using the average pairwise force, as the application of the external force changes the relative populations of attractive versus repulsive forces, shifting the pairwise forces population towards an

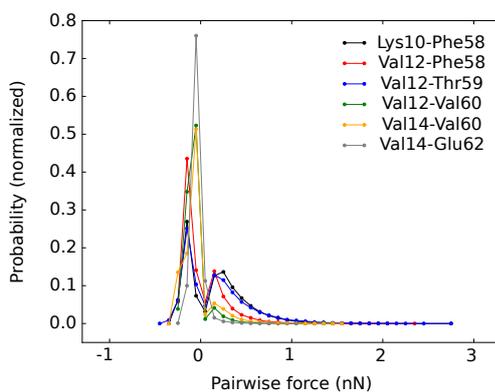
An approximately bi-modal statistical distribution, like the one in Figure 4.11a, cannot be properly characterized by the average (or arithmetic mean) value. The average can be significantly influenced by the larger positive values, even though they appear with low frequency and are thus not representative for the overall set of values; the median is a better statistical measure in such cases. The inset of Figure 4.11a compares the average and median values for the residue pairwise forces in Figure 4.11a.

For some of the H-bonds, the two measures

approximately single standard distribution (Figure 4.11b), for which the average value becomes again relevant.

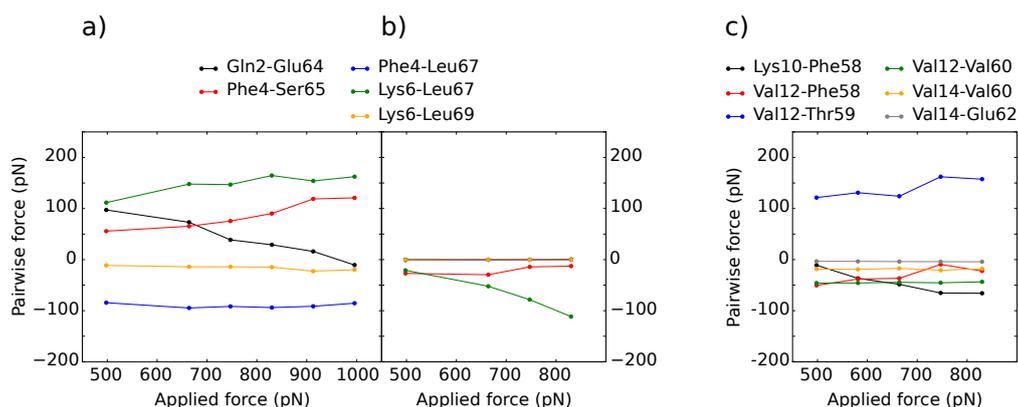
The residues in the  $\beta_1$  and  $\beta_4$  strands of NuG2 can also form H-bonds, however their probabilities to be present during equilibrium simulations vary much more than in ubiquitin (Table 4.7), suggesting a less stable arrangement. The non-backbone H-bond Val12(CO)-Thr59(OH) competes with the backbone H-bond Val12(CO)-Val60(NH), such that Val12(CO) can form H-bonds for more than 75% of the time. The fluctuations of pairwise forces between the residues forming these H-bonds in equilibrium simulations are shown in Figure 4.12. The Val14-Glu62 H-bond is located at the C-terminal edge of the  $\beta$ -strands and, for the short times it exists, it samples mostly low attractive forces, indicating a low participation in the mechanical resistance. The other H-bonds sample both negative (attractive) and positive (repulsive) values, with the negative forces appearing more often than the positive ones; this contrast to the  $\beta_1$ - $\beta_5$  H-bonds in ubiquitin, where the sampling is more balanced. Together with the lower probabilities of H-bonds presence, this suggests that interface between the central  $\beta$ -strands in NuG2 is in a slightly tensed state, even in equilibrium. This is another expression of the pre-stress found in ubiquitin<sup>41</sup>.

From the force-clamp MD simulations on ubiquitin, we determine the variation of pairwise forces between the residues forming H-bonds as a function of the applied force (Figure 4.13a), giving an indication of their relative participation in the mechanical resistance. During the overall waiting time, the structure of the protein is still close to the native state. Therefore, it is not surprising that some of the pairwise forces in H-bonds maintain their values independent of the applied force; three of them show changes correlated to a slight



**Figure 4.12:** Fluctuations of pairwise forces between residues forming H-bonds located on the  $\beta_1$  and  $\beta_4$  strands of NuG2, during equilibrium simulations.

displacement of the N-terminal ends of the two  $\beta$ -strands away from each other. The picture is quite different during the semi-stable state with end-to-end distances between 5.0 and 6.0 nm (Figure 4.13b), when most of the native contacts between the two  $\beta$ -strands are lost and some of the residues become too distant to form H-bonds. Only one of the native H-bonds (Lys6-Leu67) shows an increase in its attractive nature which correlates well with the increase in the external force, suggesting that this H-bond becomes an important element of mechanical resistance to unfolding.



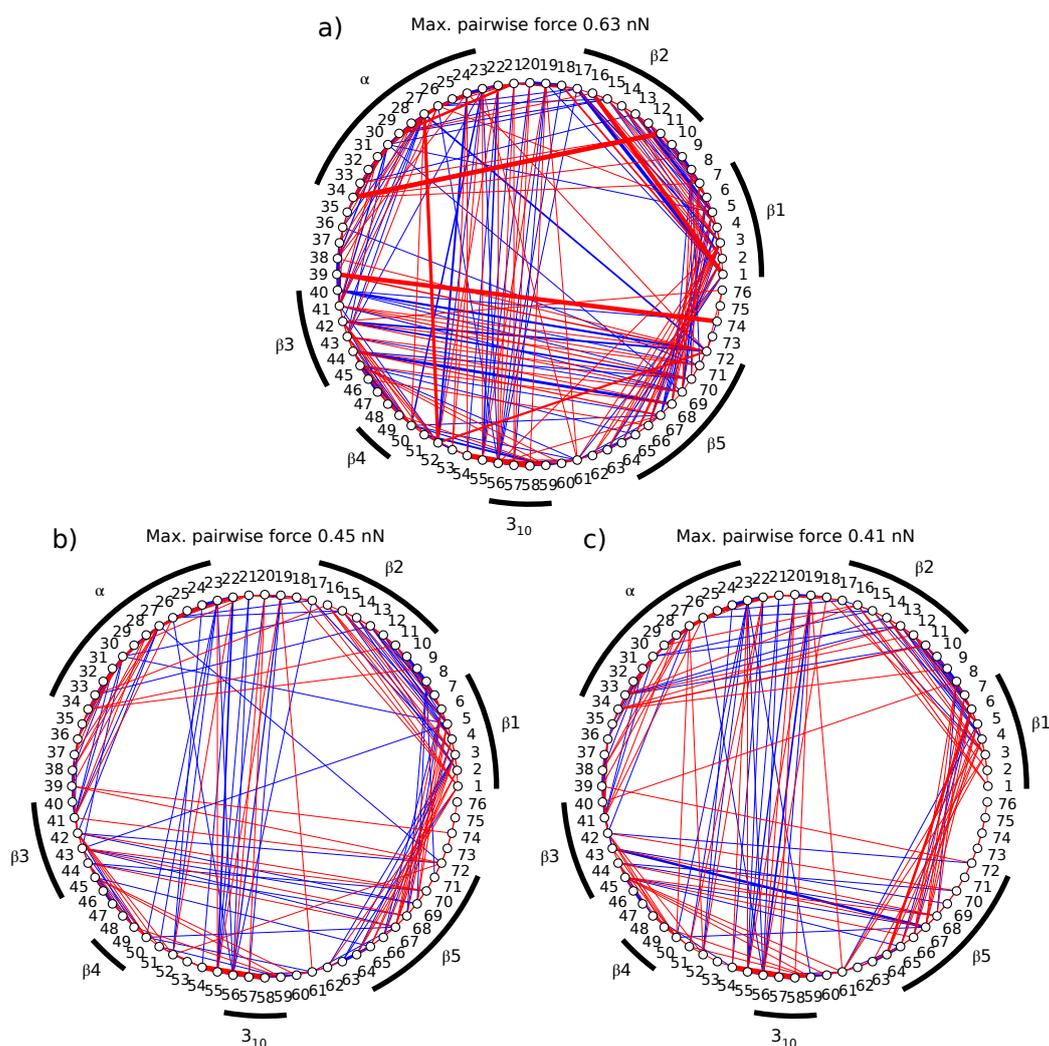
**Figure 4.13:** Average pairwise forces as a function of the applied force, from plateaus in the range of 4-5 nm (a) and 5-6 nm (b) for ubiquitin, and from plateaus in the range of 4.5-5.5 nm for NuG2 (c).

During the overall waiting time, the pairwise forces in the native H-bonds between the  $\beta_1$  and  $\beta_4$  strands of NuG2 also remain relatively independent of the applied force (Figure 4.13c), consistent with maintaining a native-like structure. The H-bonds showing correlated changes are located towards the N-terminal end of the strands, indicating a slight displacement of the strands with respect to each other as the external force increases. For NuG2, the semi-stable state could only be observed in simulations at 498 and 581 pN, and the number of samples is too low for a relevant statistical analysis of pairwise forces in the residues forming H-bonds.

The above results show that the pairwise forces in H-bonds remain relatively independent of the external force during the overall waiting time, for both ubiquitin and NuG2. This suggests that other structural elements participate in the mechanical resistance, bearing part of the load, at this stage of the unfolding process. In the next section, we try to identify these elements and provide a better understanding of how the applied force distributes through the structure of the proteins.

#### 4.6.3 Force distribution in stretched proteins

The representation of average residue pairwise forces in equilibrium (Figure 4.14a) shows a mesh of attractive and repulsive interactions, an expression of the pre-stress<sup>41</sup>. Large attractive forces correspond to electrostatic interactions between charged residues; they appear both within a secondary structure unit (in the  $\alpha$  and  $3_{10}$  helices) and between units (like  $\alpha$ - $\beta_2$ ,  $\beta_1$ - $\beta_2$ , etc.). Large repulsive forces correspond to steric clashes (like Met1-Val17) between residues brought close to each other by the adjacent attractive interactions (Met1-Glu16 and Met1-Glu18, in this case). The overall distribution of strong forces, with multiple interactions between secondary structure units, suggests a well balanced struc-



**Figure 4.14:** Average residue pairwise forces in ubiquitin from equilibrium simulations (a). Average projection of residue pairwise forces on the end-to-end distance from plateaus in the range of 4.0–5.0 nm (b) and 5.0–6.0 nm (c) obtained at 664 pN.

ture in equilibrium. The ordering of the  $\beta$ -strands in the  $\beta$ -sheet is also evident, following the chain  $\beta_4$ - $\beta_3$ - $\beta_5$ - $\beta_1$ - $\beta_2$ . The strong attractive forces involving residues from both the N- and C-terminal ends suggest a significant resistance to deformation in the end-to-end direction.

A more direct representation of the distribution of the external force inside the structure of the protein can be obtained by projecting the pairwise residue forces onto the end-to-end direction. Under the effect of the force, the floppy tail of the protein straightens and the protein is oriented such that the two application points (N- and C-terminal residues) become aligned to the force. Thus, such projections express the contribution of each pairwise residue interaction to the total force along the end-to-end direction, and also give information about their repulsive or attractive nature (represented as positive or negative values,

respectively). Figures 4.14b and c show the average projections of the pairwise forces between the  $\beta_1$  and  $\beta_5$  strands on the end-to-end direction, for the different end-to-end distance plateaus. Some differences with respect to the distribution of forces in equilibration (Figure 4.14a) are immediately visible. The highest forces in the end-to-end direction remain within and in the direct neighborhood of the  $3_{10}$ -helix, which, however, is not involved in interactions with the  $\beta_1$  and  $\beta_5$  strands, indicating that the high values of the projections occur simply due to a favorable orientation. The other interactions seem weaker than in equilibrium, however this is due to the difference in orientation, as most forces are not oriented parallel to the end-to-end direction.

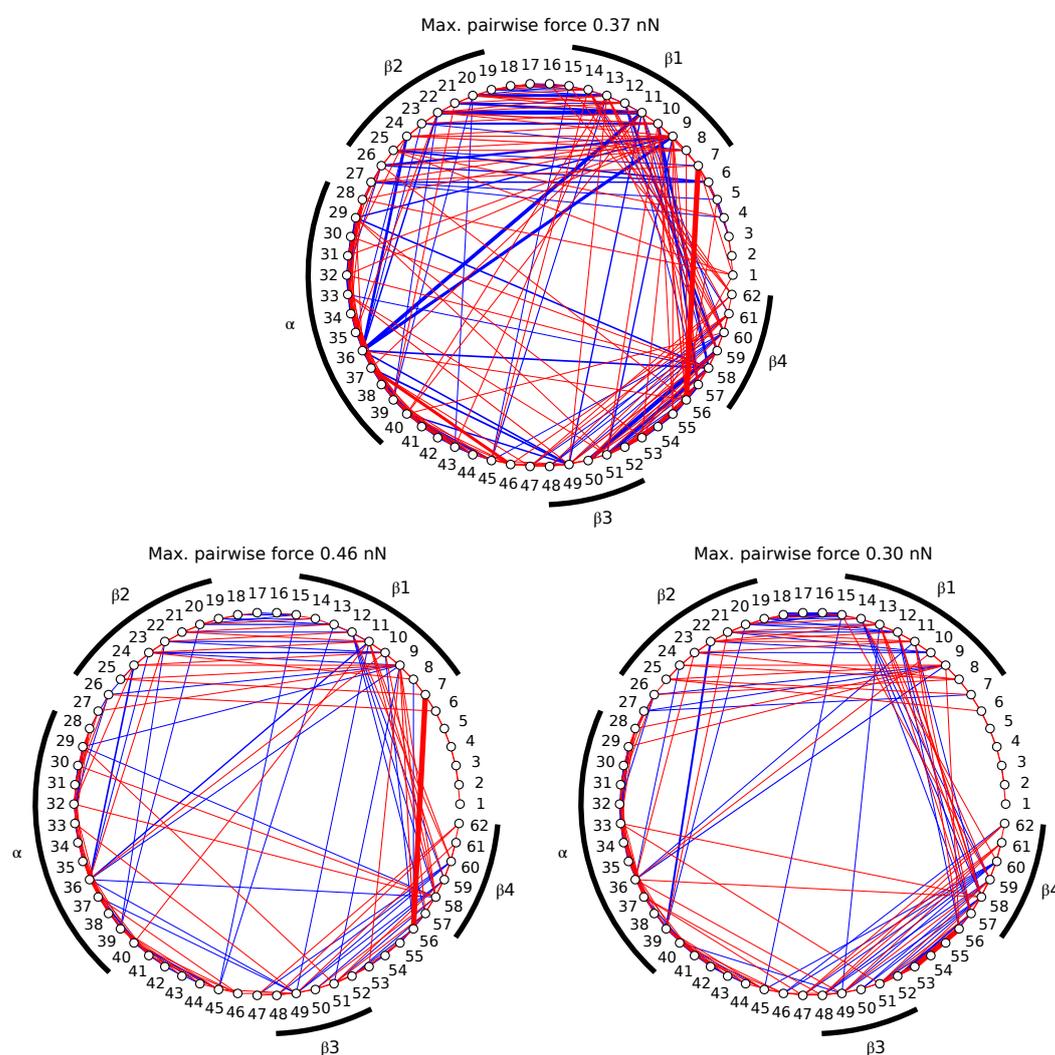
A significant difference is visible in the interactions of Asp52. In equilibrium, it has strong attractive interactions with Lys27 and Arg72, which repel each other, suggesting that these three residues are in close proximity. During the overall waiting time (Figure 4.14b), the interaction Asp52-Arg72 is much stronger than Asp52-Lys27. In the semi-stable structures with end-to-end distances in the range 5-6 nm (Figure 4.14c), the interaction Asp52-Arg72 disappears, while Asp52-Lys27 is restored. This suggests an optimization to resist deformation: when the protein is stretched but retains most of its native structure, the effort is directed towards maintaining the  $\beta_5$  strand close to the rest of the protein; when the external force overcomes the mechanical resistance and the  $\beta_5$  strand starts to slide away, the effort is redirected towards protecting the structure of the protein core. This agrees very well with the previous finding showing that the  $\alpha$  and  $\beta_3$ - $\beta_4$  hairpin resist together until late in the unfolding process<sup>92</sup>.

The sum of the average projection of pairwise forces between residues of the  $\beta_1$  and  $\beta_5$  strands on the end-to-end direction, obtained from the equilibrium simulations and at various external forces, are listed in Table 4.8. We note that there are no semi-stable structures for external forces of 913 and 996 pN. In the absence of an external force, our results show the interactions between the two  $\beta$ -strands to be very weak in the direction of the end-to-end distance. On the contrary, the two strands oppose the mechanical deformation by switching to an effective attractive interaction when the terminal residues are pulled

External force (pN)	0	498	581	664	747	830
$F_{\text{avgproj}} 4\text{-}5 \text{ nm (pN)}$	-14	-228	-271	-301	-304	-354
$F_{\text{avgproj}} 5\text{-}6 \text{ nm (pN)}$		-351	-416	-425	-582	-673

**Table 4.8:** Sums of average projections of pairwise forces onto the end-to-end direction for residues of the  $\beta_1$  and  $\beta_5$  strands in ubiquitin. For the zero force, the average was computed over the 500 equilibrium simulations used to produce the starting structures for the force-clamp simulations; the residue pairwise forces were projected onto the direction determined by the C $\alpha$  atoms of residues 1 and 72 (last residue of the  $\beta_5$  strand) to eliminate the influence of the flexible tail. For the non-zero forces, the average was computed from plateaus in the range of 4-5 nm and 5-6 nm, respectively.

apart. The attractive force increases with the external force, suggesting an elastic response of the protein and offering excellent support to our previous findings linking the stretched exponential kinetic curve to the mechanical strength of the protein. The increase is small for the native-like structures with an end-to-end distance between 4 and 5 nm, confirming the previous result that the external force is also distributed over other structural elements during this waiting time. However, the increase in  $\beta_1$ - $\beta_5$  attractive force reproduces very well the increase in external force for the semi-stable structures with an end-to-end distance between 5 and 6 nm. This suggests that the interactions between the two strands become the major component of the mechanical resistance at this stage of the unfolding process.



**Figure 4.15:** Average residue pairwise forces in NuG2 from equilibrium simulations (a). Average projection of residue pairwise forces on the end-to-end distance from plateaus in the range of 4.5–5.5 nm (b) and 5.5–6.5 nm (c) obtained at 498 pN.

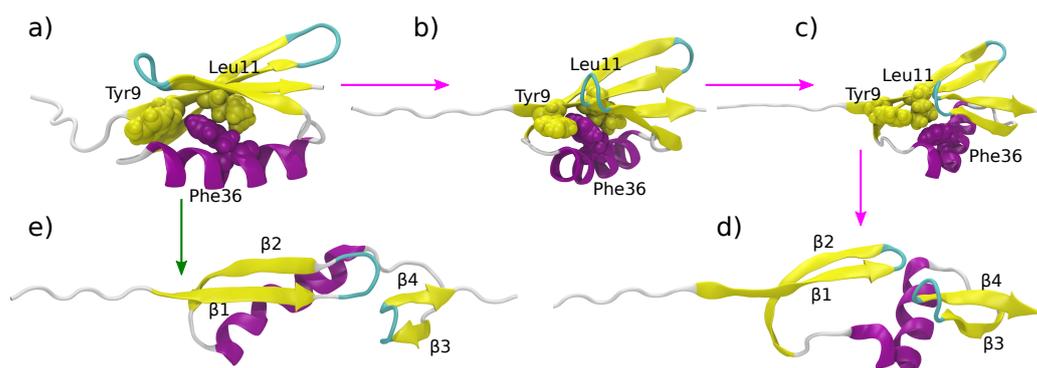
The analysis of force distribution in NuG2 reveals some interesting differences with respect to ubiquitin. The representation of average residue pairwise forces in equilibrium (Figure 4.15a) shows only a few strong attractive interactions, corresponding to salt bridges, in particular between Asp7 (part of the  $\beta_1$  strand) and Lys56 (at the edge of the  $\beta_4$  strand). More such strong attractive interactions exist in ubiquitin, but their spatial distribution is different. In ubiquitin, they hold together the different secondary structure elements, but not the two strands ( $\beta_1$  and  $\beta_5$ ) which slide apart at the start of the unfolding process. In contrast, the pairwise forces in NuG2 are mostly aligned to the primary structure, the only notable exception being the interaction between Asp7 and Lys56. This suggests the combination of H-bonds and this salt bridge as the basis of the mechanical resistance of NuG2.

Average projections of the residue pairwise forces on the end-to-end direction for the plateaus between 4.5 and 5.5 nm (Figure 4.15b) show indeed that the Asp7-Lys56 interaction is the strongest in the direction of the external force. However, this interaction completely disappears for the plateaus between 5.5 and 6.5 nm (Figure 4.15c). This is not surprising, given that Asp7 and Lys56 are both located at the edge of the respective  $\beta$ -strands, and become spatially too distant to form a salt bridge as the end-to-end distance increases. But how can the protein still resist unfolding in this case?

The sum of average projections of pairwise forces between the  $\beta_1$  and  $\beta_4$  strands for the plateaus between 4.5 and 5.5 nm (Table 4.9) varies very little with the increase in external force. Similar to ubiquitin, it is thus clear that the mechanical resistance cannot be attributed to these attractive interactions alone, and that another mechanism should oppose unfolding, both in addition to the Asp7-Lys56 salt bridge and after the salt bridge disappears. The first clue is offered by Figure 4.15a, showing strong interactions between the  $\beta_1$  and  $\beta_2$  strands, as well as between the  $\beta_3$  and  $\beta_4$  strands. These are an indicator of very stable  $\beta$ -hairpins, which can retain their structure until late in the unfolding process. The same strong forces between consecutive  $\beta$ -strands appear in Figures 4.15b and c, which represent the protein under tensile stress. In addition, the stability of the  $\alpha$ -helix is also evident, especially around its middle (Figure 4.15). The second clue is the presence

External force (pN)	0	498	581	664	747	830
$F_{\text{avgproj}} 4.5\text{-}5.5 \text{ nm}$ (pN)	-335	-548	-572	-581	-575	-593

**Table 4.9:** Sums of average projections of pairwise forces onto the end-to-end direction for residues of the  $\beta_1$  and  $\beta_4$  (including Lys56) strands in NuG2. For the zero force, the average was computed over the 500 equilibrium simulations used to produce the starting structures for the force-clamp simulations; the residue pairwise forces were projected onto the direction determined by the  $C\alpha$  atoms of residues 7 (first residue of the  $\beta_5$  strand) and 62 to eliminate the influence of the flexible region. For the non-zero forces, the average was computed from plateaus in the range of 4.5-5.5 nm.



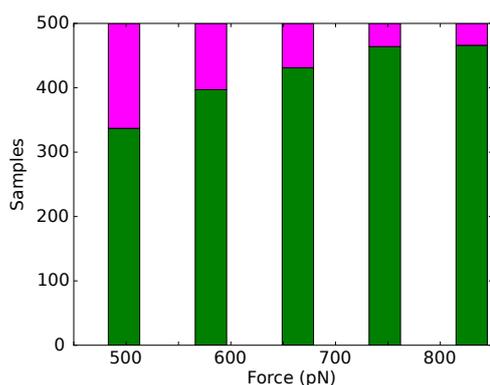
**Figure 4.16:** NuG2 unfolding. Steric clash in equilibrium (a) and under external force (b). The  $\alpha$ -helix deforms to allow Leu11 to move over Phe36 sidechain (c) during first unfolding mechanism (magenta arrows), leading to the movement of the  $\beta$ 1- $\beta$ 2 hairpin away from the protein core (d). The second unfolding mechanism (green arrow) occurs through the movement of the  $\beta$ 3- $\beta$ 4 hairpin away from the protein core, and does not feature the  $\alpha$ -helix deformation (e).

of strong repulsive interactions between residues Tyr9 and Leu11 (in the middle of the  $\beta$ 1 strand) and Phe36 (in the middle of the  $\alpha$ -helix), suggesting steric clashes between these aminoacids with relatively long side chains. Indeed, the phenyl group of Phe36 is clamped between the side chains of Tyr9 and Leu11, such that a movement of the  $\beta$ 1 chain - either in the direction of the external force or in the opposite direction - is hindered (Figure 4.16a and b). Thus, Leu11 can only be pulled by the force if the Phe36 sidechain rotates or bends, the latter being accompanied by a deformation of the  $\alpha$ -helix (Figure 4.16c). Once this occurs, the unfolding process advances quickly (Figure 4.16d).

Figure 4.15a also shows several strong interactions between the  $\alpha$ -helix and  $\beta$ 3 strand, in particular the salt bridge formed by residues Lys37 and Asp46. A weakening of these interactions allows the  $\beta$ 3 strand to move away from the  $\alpha$ -helix, once the Asp7-Lys56 bridge is broken, suggesting a different unfolding mechanism. However, the breaking of the Lys37-Asp46 salt bridge does not immediately lead to unfolding. As the  $\beta$ 3 strand and the  $\alpha$ -helix slide away from each other, other salt bridges (Asn41-Asp46 or Lys37-Glu48) can form for short periods of time. They are weak and therefore they cannot prevent, but only slow down, the further stretching of the protein. During this process, the  $\beta$ 1 strand and the  $\alpha$ -helix are kept together by the clamp formed by Tyr9 and Leu11 around Phe36 (Figure 4.16e).

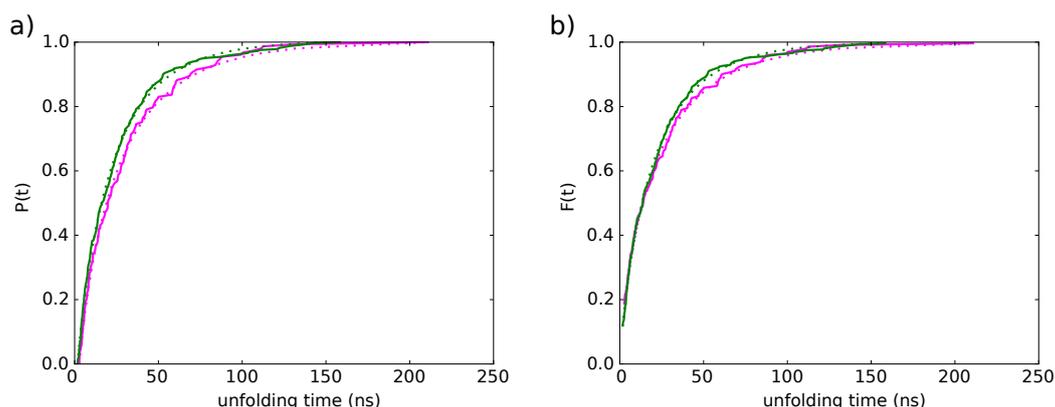
A complete picture of the mechanical unfolding of NuG2 now emerges. The protein is composed of three mechanically stable elements: two  $\beta$ -hairpins and an  $\alpha$ -helix, connected by short flexible loops. In equilibrium, a mesh of interactions exists between them, similar to the pre-stress network found in ubiquitin<sup>41</sup>. The interactions include H-bonds between the  $\beta$ 1 and  $\beta$ 4 strands, several salt bridges (most importantly Asp7-Lys56 and

Lys37-Asp46), and a clamp formed by the sidechains of Tyr9 and Leu11 surrounding the phenyl group of Phe36. Upon application of the external force, all these interactions oppose the mechanical deformation. The salt bridge Asp7-Lys56 holds the  $\beta_1$  and  $\beta_4$  strands together, is natively aligned almost parallel to the direction of the external force, and withstands large loads (around 500 pN) before yielding. The breaking of this salt bridge represents the starting event of the unfolding process, after which two mechanisms compete. The first one consists of the  $\beta_1$ - $\beta_2$  hairpin sliding away from the rest of the structure. It involves breaking of H-bonds and stronger salt bridges, and overcoming a steric clash. The second mechanism consists of the  $\beta_3$ - $\beta_4$  hairpin moving with respect to the  $\alpha$ -helix and  $\beta_1$ - $\beta_2$  hairpin, which are stabilized by the Tyr9-Leu11 clamp around Phe36. It is simpler, involving only the breaking of H-bonds and weaker salt bridges, and thus features a lower mechanical resistance to deformation (comparison of salt bridges strength based on Figure 4.15a). Taken together, this suggests that the second mechanism has a higher probability of occurrence than the first mechanism, and that this probability increases with the external force. This is indeed observed in Figure 4.17.



**Figure 4.17:** Frequency of appearance of the two unfolding mechanisms of NuG2 at different external forces. The first mechanism (the  $\beta_1$  strand moving away from the  $\alpha$ -helix) is represented in magenta, the second mechanism (the  $\beta_3$  strand moving away from the  $\alpha$ -helix) is represented in green.

In theory, the two mechanisms for the mechanical unfolding of NuG2 can lead to different unfolding kinetics. We can compare their kinetic profiles using the same procedure we applied earlier, after splitting the 500 samples that were exposed to an external force of 498 pN into two sets corresponding to the different mechanisms. We observe differences (Figure 4.18) which correlate well to the type of mechanism. The first mechanism is more complex, involving the presence of the Tyr9-Leu11 clamp around Phe36, which delays the sliding of the  $\beta_1$ -strand away from the rest of the protein and thus the unfolding process; this corresponds to a slightly more stretched exponential (a smaller exponential constant  $b=0.77$ ). In contrast, the simpler second mechanism has a kinetic profile closer to a single exponential ( $b=0.84$ ), which is the generally accepted model for processes involving breaking of a bond. At higher forces, the decreasing amount of samples showing the first mechanism leads to less reliable fitting results. The kinetic profiles for both mechanisms still show a stretched exponential behavior, such that a combination of them in any ratio will maintain the overall stretched exponential profile. This can become relevant in a comparison with experimental data, as the forces used for me-



**Figure 4.18:** Kinetic profiles from unfolding of NuG2 at 498 pN separately from first (magenta) and second (green) mechanism. Dotted lines represent stretched exponential fits. a) Unfolding profiles after elimination of points below  $t_{\min}$ . b) Modified CDF after elimination of points below  $t_{\min}$ .

chanical unfolding in AFM experiments are significantly lower than those used in MD simulations, and the preference towards one or the other mechanism was shown above to depend on the force.

From the 23 samples that show plateaus between 5.5 and 6.5 nm for an external force of 498 pN, 8 samples unfold through the first mechanism and 15 samples through the second one. In spite of the small amount of samples, this reveals a preference for the second mechanism, which agrees with the overall occurrence probabilities (Figure 4.17). It also suggests that the initial stretching of the protein is independent of the mechanism. From a kinetic point of view, this indicates that the stretched exponential unfolding is a common characteristic which occurs either in the initial phases of the unfolding, or later in both unfolding mechanisms.

#### 4.7 SUMMARY AND CONCLUSIONS

Using force-clamp MD simulations, we investigated the mechanical unfolding of two proteins with a similar secondary structure, and tried to establish a link between the non-exponential unfolding kinetics and the force or stress distribution in the structure of the proteins. We determined force-dependent unfolding rates which we combined with experimental ones in a DHS model covering one order of magnitude of applied forces. Using TRFDA, we then investigated the distribution of forces and stresses in proteins, with most of the results having relevance in the context of the unfolding kinetics.

Experimental studies<sup>3,4,14</sup> have reported non-exponential kinetics for mechanical unfolding of several proteins, and the data is best modeled by a stretched exponential (or Weibull) function<sup>84</sup>. We were able to replicate the stretched exponential kinetics for both ubiquitin

and NuG2 in our simulations, and derive kinetic parameters that agree well with experimental ones. We found that the unfolding kinetic curves become more stretched with the decreasing mechanical resistance of the protein. Our results also suggest that the two-state kinetic model of protein unfolding should be augmented by a component expressing the elastic behavior of the protein stretched by an external force. The validity of the single exponential model, assumed so far for protein unfolding under any conditions, should therefore be limited to spontaneous unfolding, in the absence of force.

The Bell and DHS model (Section 1.2) assume that the protein has a unique unfolding pathway. Several of our results suggest that multiple pathways are present in the unfolding process of ubiquitin and NuG2, supporting the theory of glassy dynamics<sup>3</sup> or static disorder in the transition states<sup>5</sup> which have been proposed as explanation for the stretched exponential kinetics. We note though that the unfolding rate constants obtained for ubiquitin from our simulations show largely different values from the experimental ones and a distinct slope, suggesting that the force-clamp MD simulations sample different regions of the energy landscape - a hypothesis for further investigation. A possible explanation is the different timescales involved: in experimental setups, the slow deformation allows the protein to be permanently in quasi-equilibrium. In contrast, the timescale of observed unfolding in our force-clamp MD simulations is close to the simulation timescale needed for proteins to reach an equilibrium state (nanoseconds to tens of nanoseconds). It is therefore conceivable that the two processes can take place simultaneously and compete, influencing the observed unfolding process and the underlying pathway. This is also supported by the existence of the semi-stable structures, with increased frequency as the applied force decreases. Still, this is among the first studies to combine and compare experimental and MD simulations kinetic data in a single model, and thus the prediction of equilibrium kinetic parameters which agree well with experimental ones is a remarkable result.

The cubic DHS model best describes the energy landscape of ubiquitin unfolding, while the cusp model seems more appropriate for NuG2. This is a surprising result, given that the two proteins share the same  $\beta$ -grasp motif and were thought to have the same unfolding mechanism based on the rupture of hydrogen bonds. This suggests that the energy landscape is dictated by the residue or even atomic level differences between the two proteins; indeed, the two strands with major role in the mechanical resistance ( $\beta_1$  and  $\beta_5$  for ubiquitin,  $\beta_1$  and  $\beta_4$  for NuG2) have a low degree of sequence similarity.

By analysis of the internal force distribution, we identified two unfolding mechanisms for NuG2. One mechanism involves breaking of hydrogen bonds and salt bridges, similar to ubiquitin; the other mechanism additionally involves a steric clash which needs to be overcome during the unfolding process. Furthermore, in NuG2, a strong salt bridge directly opposes the sliding of the two  $\beta$ -strands away from each other, while in ubiquitin only

weaker H-bonds take on this role. Also, in NuG2 the two hairpins and the  $\alpha$ -helix act like rigid bodies in the initial phases of the unfolding process; in ubiquitin, the  $\beta_5$  strand is not part of any hairpin and, as a consequence, is more flexible. All these differences indicate that the unfolding processes cannot be directly compared, despite the high secondary and tertiary structure similarity between the two proteins. In a previous study, differences were found in unfolding pathways for two other proteins of the ubiquitin family<sup>92</sup>. Thus, we conclude that mechanical strength and its underlying mechanism can significantly vary, even for proteins containing similar structural elements or motifs. This conclusion is of particular relevance for studies of mechanical resistance of models built based on structural homology: the strength of the reference molecule is not a good predictor of the strength of the model.

The distance to the transition state along the unfolding coordinate is often compared to the length of an H-bond which presumably breaks, constituting the crucial event in the unfolding process<sup>14</sup>. Our results from forced unfolding of NuG2 suggest that this is inappropriate when the exact unfolding mechanism is not known. The major initial event in the unfolding process of NuG2 is the breaking of a salt bridge, which has a different typical bond length from an H-bond. Furthermore, NuG2 has shown two unfolding mechanisms in our force-clamp simulations, with a preference for one or the other depending on the magnitude of the external force. Considering in this case that the breaking of an H-bond is the crucial unfolding event is equivalent to making one of the following assumptions: in the experimental conditions, only one of the two unfolding mechanisms takes place, or both mechanisms involve H-bond breaking as the crucial event. In the absence of detailed information about the unfolding process, making such assumptions is risky.

Furthermore, our results indicate that the distance to the transition state along the unfolding coordinate for ubiquitin is smaller than for NuG2. We also find that the H-bonds between the  $\beta_1$  and  $\beta_5$  strands in ubiquitin become the major component of the mechanical resistance once the unfolding process starts. In contrast, in NuG2, salt bridges, with a higher typical bond length than H-bonds, are involved in the mechanical resistance both in the native-like state and later in the unfolding process. This suggests that the distance to the transition state correlates with the underlying interaction which confers the mechanical resistance.

In a significant number of samples, we observed semi-stable structures. These become more frequent at lower external forces, suggesting that a high population of such structures might be encountered for the experimental range of forces. However, the difference in the end-to-end distance is apparently too small to be noticed in force spectroscopy experiments. Such semi-stable structures could potentially appear in the unfolding process of other  $\beta$ -stranded proteins. As observed for both ubiquitin and NuG2, the native struc-

ture contains several H-bonds, while the semi-stable structure contains fewer, rearranged H-bonds. As the H-bonds are often directly involved in the mechanical resilience, an intermediate with less H-bonds can only be present if the rest of the protein structure is able to dynamically compensate the loss of mechanical resistance through a different internal force distribution. This was observed for both ubiquitin and NuG2, and is thus likely to have a more general character. Investigation of the initial unfolding stages for a larger range of proteins, including some with  $\beta$ -only secondary structure, like titin or the silk protein, would be required, and should constitute the subject of further research.

The stretched exponential kinetics could be the expression of an additional state in the unfolding process. However, we did not find evidence of such additional state, and the semi-stable states that we observed were randomly distributed among the samples. We therefore conclude that the stretched exponential kinetics is simply an expression of the protein elasticity upon mechanical deformation. External forces make the folded proteins sample a broad energy landscape of native-like substates, which eventually unfold. From a kinetic point of view, the protein is not an on-off switch (or a combination thereof) as assumed in the Arrhenius equation, but has a continuous elastic response to an external force. Our findings, as well as the experimental results<sup>3,4,14</sup>, suggest that this might be a generic characteristic in the mechanical unfolding of proteins.

# 5

---

## CONCLUSIONS AND FURTHER PERSPECTIVES

---

Biomolecules and other macromolecules have long been studied with the help of MD simulations. Their results are typically expressed in terms of atomic coordinates, which are then used for further analysis. Interactions between atoms or residues are therefore only expressed indirectly, through the effects they have on the molecular structure. Force Distribution Analysis<sup>42</sup> was introduced to directly reveal the individual interactions and their spatial distribution in the simulated molecules, as found in equilibrium states. Many biomolecules undergo conformational transitions upon exposure to external perturbations, such as the binding of a ligand or the application of a mechanical force. This property is often encountered during communication processes at molecular level, for example in the allosteric transmission of signals. To follow the fluctuations of intra- and inter-molecular interactions during such dynamic processes, we developed the Time-Resolved Force Distribution Analysis (TRFDA) and implemented it as an extension of the popular MD software GROMACS<sup>43</sup>.

TRFDA is centered around the concept of pairwise forces, representing interactions between pairs of atoms or residues. It contains a force decomposition scheme for 3- and 4-body energy potentials, and can accurately represent the distribution of an external force throughout the molecular structure. It can natively calculate forces between residues, thus allowing an easy mapping of the interactions to the structure of a protein. It can also calculate a per atom or per residue punctual stress, which highlights the points where forces accumulate in the molecular structure, thereby helping to identify important elements which contribute to the mechanical resistance of the molecule.

The first application of TRFDA was the study of stress distribution during the indentation until rupture of a single layer graphene sheet. While graphene is not a biomolecule and its indentation cannot be considered a communication process, this project has provided essential insight into the distribution of stress in molecular structures during material deformation and rupture, and served as a test-case for the TRFDA development. We were able to show that the stress accumulates under the AFM indenter tip much stronger than previously assumed, and that the stress decays to almost background levels at distances as low as 5-10 times the indenter radius. This suggests that a small area of the sheet found directly under the indenter is able to bear most of the load, while the rest of the sheet offers mechanical support to this central area - a surprising finding which might apply to other non-linearly elastic materials. The analysis of forces in C-C bonds has revealed that thermal fluctuations initiate the material rupture, and that the probability of rupture decreases exponentially as the distance from the indenter tip increases, allowing us to explain the locality of the material failure.

Studies of mechanical stability of proteins can reveal many details related to their structure and function. From force-clamp MD simulations of protein unfolding, we were able to reproduce a previously reported stretched exponential kinetic profile of ubiquitin and found a similar profile for the structurally analogous NuG2. Our analysis, based mostly on TRFDA, suggests that the stretched exponential kinetics is simply an expression of the protein elasticity. The proteins do not behave like a binary switch, as implied by the widely used Arrhenius equation, but have a continuous elastic response to the external force. TRFDA revealed a dynamic distribution of interactions throughout the protein structure. The forces exist also at equilibrium<sup>41</sup>, and their variations are correlated to the progress of the unfolding process. TRFDA allowed us to identify the structural elements bearing most of the external load. Salt bridges and H-bonds oppose the mechanical deformation, and they gradually yield upon prolonged application of force. Some of the secondary structure units, like  $\beta$ -hairpins and  $\alpha$ -helices, are much more stable than others, acting similar to rigid bodies and therefore influencing the unfolding process. TRFDA also helped uncover that the mechanical unfolding of NuG2 can happen through two different mechanisms.

Their relative probability of occurrence seems to be dependent on the external force, such that, at the large forces used in force-clamp MD simulations, one mechanism is observed frequently, while the other mechanism is more probable in the range of lower experimental forces. Thus, a direct link between the structural elements and the unfolding process cannot be made: proteins with very similar secondary and tertiary structures, like ubiquitin and NuG2, may have different unfolding mechanisms and different mechanical properties.

Pairwise forces have an inherently noisy nature in most biomolecular systems, mainly due to the thermal fluctuations of the atoms. The statistical distribution of their values can be bi-modal in equilibrium simulations, such that the arithmetic mean and standard deviation (or variance), as previously obtained from FDA<sup>42</sup>, might not be the best way of representing the typical values. This suggests that a more thorough statistical analysis of the variation of pairwise forces should be carried out, in order to determine the best parameters to describe them in different types of simulations. To express the overall character of the interactions between two atoms or residues with a single value, more robust statistical measures like the median or a weighted average might be more appropriate than the average.

The noisy nature of the pairwise forces also raises difficulties in analyzing them. It is often more interesting to find the trend in the variation of the pairwise forces, rather than the amount they fluctuate. For example, such a trend could express the overall repulsive character of a compressed bond, or the amount of attractive force that opposes an external mechanical load. Averaging pairwise forces over time, space or molecular ensembles represents an effective method of noise reduction. Typical signal denoising techniques, like filtering, methods based on Fourier transforms, or total variation, can also be applied. In all these cases, however, the statistical implications of the noise reduction method should be carefully considered. This is another research avenue that should be followed.

Defining stress at atomic level is not straightforward. Stress is typically defined as the ratio between force and the surface on which it acts. While forces are a natural component in MD simulations, it is difficult to define atomic level areas. The local pressure concept<sup>48</sup> and the circular stress, introduced in this work, use geometrical constructs which are fixed in space and on which an area can be easily defined. The atomic virial stress<sup>47</sup> uses a different approach, being expressed as the ratio between energy and volume, which is equivalent to the true stress definition; however, atomic level volumes are also not well defined, so the atomic virial stress is typically measured in energy units. Similarly, the punctual stress, also introduced in this work, is expressed as a sum of the absolute values of pairwise forces, as the area of action of pairwise forces is ill-defined. The comparison of punctual, circular and virial stress in graphene revealed that all of them are useful measures of the accumulation of forces in molecular systems, being able to predict well the location of material failure.

The redistribution of stress in the indented graphene is almost instantaneous, while, in a protein, the stress levels change within picoseconds after applying an external load. This indicates that the force propagation is very fast in a wide range of molecular structures. In proteins, this contrasts strongly to unfolding times of nanoseconds or larger, showing how such biomaterials have evolved to resist mechanical force - the simple presence of force is not enough to destroy the structure, a prolonged application is required to dismantle the native contacts. When averaged over an ensemble, proteins in equilibrium appear to have a constant level of mechanical stress; the same is true during the waiting time in an unfolding process, suggesting that the stress can be used to define (quasi-)stable states of the ensemble between which the protein transitions fast. The proteins used in this work have well defined structures - but can the same concept be applied to natively disordered proteins, in order to cluster conformations with similar mechanical properties? This is another intriguing idea that should be investigated in a future project.

Even though the current experimental means do not allow a direct assessment of pairwise forces or stress within a molecular structure, we believe that these observables are very useful in analyzing and understanding the mechanical response of a complex molecule such as a protein, just as stress calculations in structural mechanics analyses of macroscopic objects proved valuable in the design process. With the advent of instruments able to apply or measure forces on single molecules, like force spectroscopy or optical tweezers, it is increasingly likely that mechanical properties at sub-molecular level (domain, motif or even secondary structure in proteins) will soon be investigated. This is a scale at which experiments would be perfectly complemented by MD simulations, which have already provided unprecedented insights into the mechanical stability and rupture mechanisms of proteins. TRFDA represents the ideal tool for analyzing such results, revealing the underlying molecular basis of the observed mechanical response.

The internal distribution of forces was already identified as the basis of allosteric signaling in the methionine repressor MetJ<sup>39</sup>. TRFDA can be used for the study of similar communication mechanisms, with the added advantage of providing a dynamic view of the signal propagation through the protein core. It could be used, for example, to study the transduction of signals from the extracellular matrix by integrins or other cell adhesion molecules which are involved in the attachment of a cell to its surroundings and in the sensing of the extracellular environment. Another possible application is the investigation of stress distribution in the cytoskeleton, in particular the interplay between microfilaments, intermediate filaments and microtubules. More dynamical processes, like those associated to cell motility, would also constitute a good subject for TRFDA.

As can be seen from the applications in the present work, as well as from the list of possible future research directions, TRFDA is a very versatile method, giving insights into

both the structure and the functionality of biomolecules. We hope that it will soon become a common tool for analyzing results from MD simulations of (bio)materials.



# A

---

## MULTIBODY FORCE DECOMPOSITION

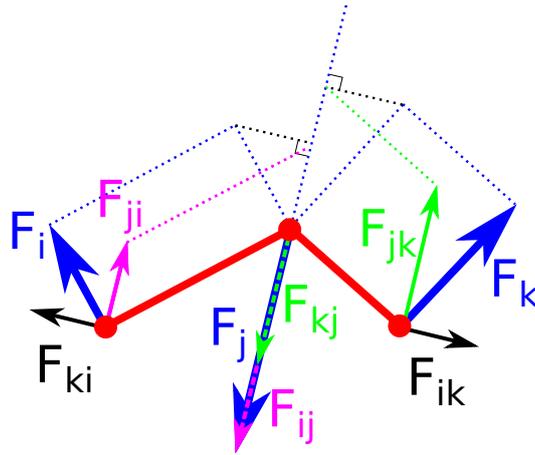
---

This appendix describes the force decomposition for 3- and 4-body potentials, which we introduced as part of TRFDA.

### A.1 FORCE DECOMPOSITION FOR 3-BODY POTENTIALS

For an angle formed by atoms  $i$ ,  $j$ ,  $k$  (forming bonds  $i$ - $j$  and  $j$ - $k$ , see Figure A.1), the sum of the atomic forces is zero:

$$\mathbf{F}_i + \mathbf{F}_j + \mathbf{F}_k = 0 \tag{A.1}$$



**Figure A.1:** Force decomposition for an angle potential.  $\mathbf{F}_i$  and  $\mathbf{F}_k$  are decomposed into a component in the direction opposite to  $\mathbf{F}_j$  and a component in the perpendicular direction.

The atomic forces  $\mathbf{F}_i$  and  $\mathbf{F}_k$  can be decomposed into a component in the direction of  $\mathbf{F}_j$  and a perpendicular component. The perpendicular components cancel each other out because of Equation A.1. The components in the direction of  $\mathbf{F}_j$  can be written as:

$$\mathbf{F}_{ij} = -\mathbf{F}_{ji} = F_i \cos(\mathbf{F}_i, -\mathbf{F}_j) \cdot \mathbf{U}_j \quad (\text{A.2})$$

and

$$\mathbf{F}_{kj} = -\mathbf{F}_{jk} = F_k \cos(\mathbf{F}_k, -\mathbf{F}_j) \cdot \mathbf{U}_j \quad (\text{A.3})$$

The pairwise force between atoms i and k can be obtained by a vector difference:

$$\mathbf{F}_{ki} = \mathbf{F}_i - \mathbf{F}_{ji} = -\mathbf{F}_{ik} = -(\mathbf{F}_k - \mathbf{F}_{jk}) \quad (\text{A.4})$$

## A.2 FORCE DECOMPOSITION FOR 4-BODY POTENTIALS

For a dihedral angle formed by atoms i, j, k, l (forming bonds i-j, j-k and k-l, see Figure A.2), the sum of the atomic forces is zero:

$$\mathbf{F}_i + \mathbf{F}_j + \mathbf{F}_k + \mathbf{F}_l = 0 \quad (\text{A.5})$$

which can also be written as:

$$\mathbf{F}_i + \mathbf{F}_l = -(\mathbf{F}_j + \mathbf{F}_k) \quad (\text{A.6})$$

Atoms  $j$  and  $k$  can be considered to form a single body on which a combined force:

$$\mathbf{F}_{j+k} = \mathbf{F}_j + \mathbf{F}_k \quad (\text{A.7})$$

acts as shown in Figure A.2a. Similar to the force decomposition for an angle,  $\mathbf{F}_i$  and  $\mathbf{F}_l$  can be decomposed in a component in the direction of  $\mathbf{F}_{j+k}$  and a component in the perpendicular direction. The components in the perpendicular direction cancel each other out because of Equation A.6. The components in the direction of  $\mathbf{F}_{j+k}$  can be written as:

$$\mathbf{F}_{i(j+k)} = F_i \cos(-\mathbf{F}_i, \mathbf{F}_{j+k}) \cdot \mathbf{U}_{j+k} \quad (\text{A.8})$$

and

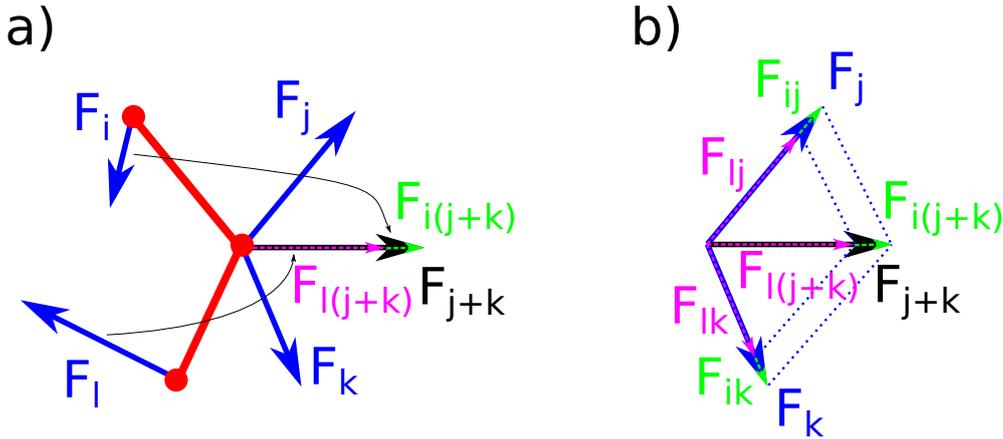
$$\mathbf{F}_{l(j+k)} = F_l \cos(-\mathbf{F}_l, \mathbf{F}_{j+k}) \cdot \mathbf{U}_{j+k} \quad (\text{A.9})$$

The pairwise forces acting on atoms  $j$  and  $k$  can be written in a generic way as:

$$\mathbf{F}_j = \mathbf{F}_{ij} + \mathbf{F}_{kj} + \mathbf{F}_{lj} \quad (\text{A.10})$$

and

$$\mathbf{F}_k = \mathbf{F}_{ik} + \mathbf{F}_{jk} + \mathbf{F}_{lk} \quad (\text{A.11})$$



**Figure A.2:** Force decomposition for a dihedral angle potential. a) Atoms  $j$  and  $k$  are considered a single body on which a combined  $\mathbf{F}_{j+k}$  acts. Forces are then decomposed similar to Figure A.1. b) The pairwise forces between the atoms  $i$  and  $l$  and the  $j+k$  single body are decomposed into components in the directions of  $\mathbf{F}_j$  and  $\mathbf{F}_k$ .

and, because  $j+k$  is considered a single body,  $j$  and  $k$  do not move with respect to each other, so the pairwise force between them is zero:

$$\mathbf{F}_{kj} = -\mathbf{F}_{jk} = 0 \quad (\text{A.12})$$

such that:

$$\mathbf{F}_{j+k} = \mathbf{F}_{ij} + \mathbf{F}_{lj} + \mathbf{F}_{ik} + \mathbf{F}_{lk} \quad (\text{A.13})$$

A decomposition of  $\mathbf{F}_{i(j+k)}$  and  $\mathbf{F}_{l(j+k)}$  can be made in the direction of  $\mathbf{F}_j$  and  $\mathbf{F}_k$ , as shown in Figure A.2b, such that:

$$\mathbf{F}_{i(j+k)} = \mathbf{F}_{ij} + \mathbf{F}_{ik} \quad (\text{A.14})$$

and

$$\mathbf{F}_{l(j+k)} = \mathbf{F}_{lj} + \mathbf{F}_{lk} \quad (\text{A.15})$$

Defining  $\alpha$  as the angle between  $\mathbf{F}_{j+k}$  and  $\mathbf{F}_j$  and  $\beta$  as the angle between  $\mathbf{F}_{j+k}$  and  $\mathbf{F}_k$ , these vectors can be written as:

$$\mathbf{F}_{ij} = -\mathbf{F}_{ji} = \frac{F_{i(j+k)} \sin \beta}{\sin \alpha \cos \beta + \sin \beta \cos \alpha} \cdot \mathbf{U}_j \quad (\text{A.16})$$

$$\mathbf{F}_{lj} = -\mathbf{F}_{jl} = \frac{F_{l(j+k)} \sin \beta}{\sin \alpha \cos \beta + \sin \beta \cos \alpha} \cdot \mathbf{U}_j \quad (\text{A.17})$$

$$\mathbf{F}_{ik} = -\mathbf{F}_{ki} = \frac{F_{i(j+k)} \sin \alpha}{\sin \alpha \cos \beta + \sin \beta \cos \alpha} \cdot \mathbf{U}_k \quad (\text{A.18})$$

$$\mathbf{F}_{lk} = -\mathbf{F}_{kl} = \frac{F_{l(j+k)} \sin \alpha}{\sin \alpha \cos \beta + \sin \beta \cos \alpha} \cdot \mathbf{U}_k \quad (\text{A.19})$$

Finally, the pairwise force between atoms  $i$  and  $l$  can be obtained by a vector difference:

$$\mathbf{F}_{li} = \mathbf{F}_i - \mathbf{F}_{ji} - \mathbf{F}_{ki} = -\mathbf{F}_{il} = -(\mathbf{F}_l - \mathbf{F}_{jl} - \mathbf{F}_{kl}) \quad (\text{A.20})$$

# B

---

## GRAPHENE RUPTURE - METHODS

---

This appendix describes the simulation setup and parameters used for the study of graphene indentation until rupture.

We employed MD simulations to study the deformation of finite graphene sheets of circular shape under spherical AFM indenters in vacuum, during which we recorded load-displacement profiles. The calculations were performed with GROMACS<sup>43</sup> 4.5.3 using the truncated Morse potential and LAMMPS<sup>76</sup> version 17Feb2012 using the AIREBO potential<sup>57</sup>. We noticed that the energy conservation was not maintained when the GROMACS calculations were performed in single precision for the larger graphene sheets. Switching to double precision restored the energy conservation, but made the calculations slower. For consistency, we ran all GROMACS calculations in double precision, including those for smaller molecular systems.

The AFM indenter was simulated as a hollow sphere generated from Argon atoms held together by pairwise harmonic potentials. The sphere was constructed from planar circu-

lar slices of different diameters, with 0.14 nm space between them. Each slice is formed by atoms placed at 0.14 nm from each other; only the distance between the first and last atom in the order of placement in one circle is different, to account for the different circle diameters. We found that a single-walled sphere is not able to sustain in some cases the mechanical stress to which it is exposed, leading to deformation and inaccurate load-displacement profiles. Therefore, the spheres used in all simulations were double-walled, the inner wall being generated as another sphere with the radius reduced by 0.14 nm. Pairwise harmonic potentials held together any two Ar atoms located within a distance of 0.25 nm from each other. This distance is smaller than double the initial distance between Ar atoms, such that harmonic potentials only exist between an atom and its direct neighbors from the same wall or from a different wall. The harmonic potentials have an equilibrium distance of 0.14 nm and a force constant of  $7 \cdot 10^6$  kJ mol<sup>-1</sup> nm<sup>-2</sup>. The force constant was chosen about one order of magnitude larger than any similar value in the OPLS-AA force field<sup>25</sup>, and a further increase resulted in instabilities of the MD simulations.

Each planar graphene sheet was generated such that the distance between bonded atoms was equal to the C-C equilibrium bond length (0.14 nm). The carbon atoms located at the edge approximated a circle, and were saturated with hydrogen atoms. The graphene sheet and the indenter sphere interacted only through a Lennard-Jones potential. A weak attractive effect of the Lennard-Jones potential was observed at small distances between the sheet and sphere.

The largest sheet simulated for this work, with a radius of 100 nm, contained over 1.2 million carbon atoms. Modeling the largest sphere, with a radius of 27.5 nm as in experiments<sup>1</sup>, required almost 1 million Ar atoms. Obtaining results for these large systems in reasonable amounts of time was only possible due to the computational simplicity of the truncated Morse potential, due to our approach of using harmonic potentials only between neighboring atoms of the sphere, and due to the high efficiency of the GROMACS code. Still, memory requirements made impossible the simulation of graphene sheets of experimental size (500 nm radius), containing over 30 million atoms.

An energy minimization using the conjugate gradient method was carried out on each molecular system until the maximum atomic force was below 10 kJ mol<sup>-1</sup> nm<sup>-1</sup>. A 1 ns equilibrium MD simulation was then performed starting with random atom velocities generated from a Boltzmann distribution corresponding to a temperature of 300 K. The temperature was maintained at 300 K by separate coupling of the graphene sheet and sphere to velocity rescaling thermostats<sup>97</sup> with time constants of 100 fs and 20 fs, respectively. The integration time step was 1 fs for GROMACS calculations and 0.5 fs for LAMMPS calculations; the lower value of 0.5 fs is commonly used with the AIREBO potential and increasing it to 1 fs led to instabilities in the MD simulations. No charges were assigned to the atoms.

Lennard-Jones interactions were calculated up to a cutoff of 1 nm. For the AIREBO potential, the torsion term was enabled and the Lennard-Jones scale factor was set to 3, leading to a cutoff of 1.02 nm. During the equilibrium MD simulations, ripples formed throughout each graphene sheet due to the thermal motions of the atoms, of amplitude and wave lengths which were in agreement with experiments<sup>61</sup>. From each resulting equilibrium trajectory, we picked the latest frame for which the average position of the C atoms in the Z direction lay within 0.05 nm from the original plane of the graphene sheet, and used it as starting structure for the indentation MD simulations. To prevent significant deformations of a graphene sheet during these equilibrium simulations due to the ripples, the C atoms at the edge of the sheet were only allowed to move in the XY plane, while the rest of the sheet was allowed to move freely in all three directions.

The equilibrium simulations were followed by constant velocity indentation simulations, with various velocities, as indicated elsewhere in the text, and a force constant of 10000 kJ mol<sup>-1</sup> nm<sup>-2</sup>. For an indentation velocity of 0.01 nm ps<sup>-1</sup>, we also performed simulations with a force constant of one order of magnitude smaller and one order of magnitude larger, without observing any significant differences.

During the indentation simulations, C atoms at the edge of the sheet were fixed. This is similar to the experimental setup where it is assumed that the graphene sheet does not slip on top of the rigid support. The sphere was placed above the graphene sheet and the indentation force acted on the center of mass of the sphere, moving it initially towards the sheet and then pressing it onto the sheet. The initial distance between the center of the sphere and the sheet was chosen such that at least 1 ns would pass before the lower side of the sphere touched the graphene sheet. In the AFM experiments, the spherical tip is fixed on the cantilever and pushed vertically into the suspended graphene sheet. To prevent lateral and rotational motion of the sphere in our MD simulations during the out-of-plane indentation along Z, three atoms of the sphere – one located at the top of the sphere and one on each side – were restricted to only move in Z direction. Restricting all sphere atoms to only move vertically would interfere with the temperature coupling, leading to large variations in temperature.

The indentation simulations using the truncated Morse potential were carried out until the first C-C bond broke, which we defined as the initial event of graphene rupture. For the AIREBO potential, the rupture was defined by C atoms moving further than 0.2 nm apart, representing the transition distance between the reactive (REBO) and Lennard-Jones components of the potential.

For each simulation, a finite graphene sheet was placed in an empty simulation box using periodic boundary conditions (PBC), leaving at least 10 nm between the graphene sheet and the box boundaries in the XY plane. In the Z direction, the PBC box was made

sufficiently large to fit the initial distance between the sphere and the graphene sheet as well as enough space to allow deformation of the sheet; furthermore, the Z dimension was at least twice as big as the sphere movement in the Z direction. Using a periodic molecular system allowed us to make use of the efficient parallelization scheme based on domain decomposition available in GROMACS 4.0 and later versions. The computation speed was not influenced by the box size as the simulations were performed in vacuum. The LAMMPS simulations did not use PBC.

In all cases, the sphere was allowed to reach the desired constant velocity before touching the graphene sheet; larger spheres contained more atoms and therefore required a higher distance from the sheet in order to reach a constant velocity. Bringing the sphere to a constant velocity of  $1 \text{ nm ps}^{-1}$  could not be achieved within the maximum distance allowed by the simulation PBC box, setting an upper limit on the indentation velocity. The simulation with the lowest velocity ( $0.00003 \text{ nm ps}^{-1}$ ) took around 21000 hours on modern CPU cores.

We also performed reverse load simulations with a modified version of GROMACS in which the pull code was changed to allow specifying a constant acceleration. The sphere movement had 3 phases: an initial phase in which it moved with constant velocity; a second phase during which a negative acceleration was set, allowing the sphere to slow down, to reach a zero velocity, and to start an accelerated movement in the opposite direction; and the final phase in which it again moved with constant velocity (Figure 3.6b).

# C

---

## PROTEIN UNFOLDING - METHODS

---

This appendix describes the simulation setup and analysis employed for the study of mechanical protein unfolding.

All MD simulations in this work were performed using GROMACS<sup>43</sup> 4.5.3 and the OPLS-AA<sup>25</sup> force field. The starting structure for ubiquitin was the Protein Data Bank entry 1UBQ<sup>87</sup>. For NuG2, we used the Protein Data Bank structure 1MI0<sup>86</sup>, which contains two protein chains. We chose the second chain and built the missing imidazole rings in the first three His residues with the Molefacture plugin of VMD<sup>44</sup>. The position of the rings and the angles  $C_{\alpha}$ - $C_{\beta}$ - $C_{\gamma}$  were not optimized in any way, as they are part of a flexible region (residues 1-6), which could easily change conformation during the equilibrium MD simulations.

After removing water molecules found in the crystal structure, each protein was solvated in a box of SPCE<sup>98</sup> water with a dodecahedral shape, with a distance of at least 1 nm between the solute and the sides of the box; the boxes have sides 6.5 nm and 7.6 nm long,

for ubiquitin and NuG2, respectively. Na<sup>+</sup> and Cl<sup>-</sup> ions were then added to obtain a 0.1 M salt concentration.

A steepest descent minimization was carried out until the maximum atomic force was below 1000 kJ/mol nm. The molecular system was then equilibrated with a 0.1 ns NVT simulation, followed by a 1 ns NpT simulation. During both of these procedures, the heavy atoms of the protein were subjected to position restraints of 1000 kJ/mol nm. The short-range neighbor list, electrostatic and van der Waals cutoffs were set to 1 nm; long-range electrostatic interactions were computed with the Particle Mesh Ewald method<sup>30,31</sup> using a FFT grid spacing of 0.16 nm. All bonds were constrained using the LINCS algorithm<sup>27,28</sup>, allowing an integration time step of 2 fs.

At the beginning of the NVT simulation, random velocities were generated based on a Boltzmann distribution corresponding to a temperature of 300 K; the temperature was then maintained at 300 K by separate coupling of protein and of water plus ions to a velocity rescaling thermostat<sup>97</sup> with a time constant of 0.1 ps. During the NpT simulation, the molecular system was coupled to an isotropic Parrinello-Rahman barostat<sup>99,100</sup> with a time constant of 2 ps and a reference pressure of 1 bar.

The structure obtained at the end of the NpT simulation and random velocities assigned to the atoms were used to start 10 new equilibrium simulations, using the same setup as the NpT simulation and a length of 100 ns. From each of these simulations, atom coordinates were saved at intervals of 2 ns, resulting in a total of 500 structures. Each of these structures was stripped of water and ions, and the protein was oriented such that the end-to-end distance vector (determined by the C<sub>α</sub> atoms of N- and C-terminal residues) was aligned to the X direction. For each protein, a new box of SPCE water molecules was built, with a triclinic shape and dimensions of 12x5x5 nm for ubiquitin, and 10x5x5 nm for NuG2. Na<sup>+</sup> and Cl<sup>-</sup> ions were then added to obtain a 0.1 M salt concentration. The same protocol for equilibration (energy minimization, followed 0.1 ns NVT and 1 ns NpT simulations) was applied to each of the 500 structures.

To study protein unfolding, each of the 500 structures and the associated atom velocities was used as starting point for force-clamp MD simulations, using the same setup as above except that only the bonds between heavy atoms and hydrogen atoms were constrained to their equilibrium values. Constant forces were applied to the C<sub>α</sub> atoms of the C- and N-terminal residues in the X direction, corresponding to the largest dimension of the water box. End-to-end distances ( $d_{e2e}$ ) were measured between the C<sub>α</sub> atoms of the terminal residues, and simulations were stopped when  $d_{e2e}$  exceeded 10 nm for ubiquitin, and 8 nm for NuG2. The forces had values of 498, 581, 664, 747, 830, 913 and 996 pN for ubiquitin, and 498, 581, 664, 747 and 830 pN for NuG2. We note that each of the 500 structures had only one set of atom velocities assigned for all force-clamp simulations. The total time for all force-

clamp MD simulations was around 80 microseconds for ubiquitin, and 20 microseconds for NuG2.

For each of the 500 samples, the trajectories obtained in the last NpT simulations and the force-clamp simulations were subjected to Time-Resolved Force Distribution Analysis, to obtain per residue punctual stress and pairwise forces.

$d_{e2e}$  forms plateaus, during which the values fluctuate. To distinguish these plateaus from the surrounding regions of increasing  $d_{e2e}$ , we performed a smoothing of the  $d_{e2e}$  time series, after which we determined the inflection points by numerical differentiation. For smoothing, we employed the Total Variation method<sup>101</sup>, typically used for signal and image denoising; this method attempts to find an ideal signal which describes the given time series data while minimizing the length of the signal.

For visualization of molecular structures, we used VMD<sup>44</sup>. Data analysis and visualization were performed with the SciPy<sup>102</sup>, NumPy<sup>103</sup> and matplotlib<sup>104</sup> Python libraries.



---

# BIBLIOGRAPHY

---

- [1] C. Lee, X. Wei, J. W. Kysar, and J. Hone, “Measurement of the Elastic Properties and Intrinsic Strength of Monolayer Graphene,” *Science*, vol. 321, no. 5887, pp. 385–388, 2008.
- [2] N. Bhatia and W. Nachbar, “Finite indentation of an elastic membrane by a spherical indenter,” *International Journal of Non-Linear Mechanics*, vol. 3, no. 3, pp. 307 – 324, 1968.
- [3] J. Brujić, R. I. Hermans Z., K. A. Walther, and J. M. Fernández, “Single-molecule force spectroscopy reveals signatures of glassy dynamics in the energy landscape of ubiquitin,” *Nat Phys*, pp. 282–286, Apr. 2006.
- [4] S. Garcia-Manyes, J. Brujić, C. L. Badilla, and J. M. Fernández, “Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin.,” *Biophys J*, vol. 93, pp. 2436–2446, Oct 2007.
- [5] T.-L. Kuo, S. Garcia-Manyes, J. Li, I. Barel, H. Lu, B. J. Berne, M. Urbakh, J. Klafter, and J. M. Fernández, “Probing static disorder in arrhenius kinetics by single-molecule force spectroscopy,” *Proc Natl Acad Sci U S A*, vol. 107, pp. 11336–11340, Jun 2010.
- [6] G. Bell, “Models for the specific adhesion of cells to cells,” *Science*, vol. 200, no. 4342, pp. 618–627, 1978.
- [7] O. K. Dudko, G. Hummer, and A. Szabo, “Intrinsic Rates and Activation Free Energies from Single-Molecule Pulling Experiments,” *Phys. Rev. Lett.*, vol. 96, p. 108101, Mar 2006.
- [8] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*. Garland Science, 5 ed., Nov. 2007.
- [9] J. R. Forman and J. Clarke, “Mechanical unfolding of proteins: insights into biology, structure and folding,” *Curr Opin Struct Biol*, vol. 17, pp. 58–66, Feb 2007.

## BIBLIOGRAPHY

- [10] E. Evans, K. Ritchie, and R. Merkel, "Sensitive force technique to probe molecular adhesion and structural linkages at biological interfaces.," *Biophys J*, vol. 68, pp. 2580–2587, Jun 1995.
- [11] M. Carrion-Vazquez, H. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernández, "The mechanical stability of ubiquitin is linkage dependent.," *Nat Struct Biol*, vol. 10, pp. 738–743, Sep 2003.
- [12] O. K. Dudko, G. Hummer, and A. Szabo, "Theory, analysis, and interpretation of single-molecule force spectroscopy experiments," *Proceedings of the National Academy of Sciences*, vol. 105, no. 41, pp. 15755–15760, 2008.
- [13] M. Schlierf, H. Li, and J. M. Fernández, "The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, pp. 7299–7304, May 2004.
- [14] Y. Cao, R. Kuske, and H. Li, "Direct observation of markovian behavior of the mechanical unfolding of individual proteins.," *Biophys J*, vol. 95, pp. 782–788, Jul 2008.
- [15] B. J. Alder and T. E. Wainwright, "Phase Transition for a Hard Sphere System," *The Journal of Chemical Physics*, vol. 27, no. 5, pp. 1208–1209, 1957.
- [16] B. J. Alder and T. E. Wainwright, "Studies in Molecular Dynamics. I. General Method," *The Journal of Chemical Physics*, vol. 31, no. 2, pp. 459–466, 1959.
- [17] A. Rahman, "Correlations in the Motion of Atoms in Liquid Argon," *Phys. Rev.*, vol. 136, pp. A405–A411, Oct 1964.
- [18] F. H. Stillinger and A. Rahman, "Improved simulation of liquid water by molecular dynamics," *The Journal of Chemical Physics*, vol. 60, no. 4, pp. 1545–1557, 1974.
- [19] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins," *Nature*, vol. 267, pp. 585–590, June 1977.
- [20] J. D. Durrant and J. A. McCammon, "Molecular dynamics simulations and drug discovery.," *BMC Biol*, vol. 9, p. 71, 2011.
- [21] S. Ferrer, J. Ruiz-Pernía, S. Martí, V. Moliner, I. Tuñón, J. Bertrán, and J. Andrés, "Hybrid schemes based on quantum mechanics/molecular mechanics simulations goals to success, problems, and perspectives.," *Adv Protein Chem Struct Biol*, vol. 85, pp. 81–142, 2011.
- [22] S. G. Estácio, "In silico strategies toward enzyme function and dynamics.," *Adv Protein Chem Struct Biol*, vol. 87, pp. 249–292, 2012.

- [23] H. A. Scheraga, M. Khalili, and A. Liwo, "Protein-folding dynamics: overview of molecular simulation techniques.," *Annu Rev Phys Chem*, vol. 58, pp. 57–83, 2007.
- [24] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, "How fast-folding proteins fold.," *Science*, vol. 334, pp. 517–520, Oct 2011.
- [25] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids," *Journal of the American Chemical Society*, vol. 118, no. 45, pp. 11225–11236, 1996.
- [26] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *Journal of Computational Physics*, vol. 23, no. 3, pp. 327 – 341, 1977.
- [27] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, "LINCS: A linear constraint solver for molecular simulations," *Journal of Computational Chemistry*, vol. 18, no. 12, pp. 1463–1472, 1997.
- [28] B. Hess, "P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation," *Journal of Chemical Theory and Computation*, vol. 4, no. 1, pp. 116–122, 2008.
- [29] P. P. Ewald, "Die Berechnung optischer und elektrostatischer Gitterpotentiale," *Annalen der Physik*, vol. 369, no. 3, pp. 253–287, 1921.
- [30] T. Darden, D. York, and L. Pedersen, "Particle Mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems," *The Journal of Chemical Physics*, vol. 98, no. 12, pp. 10089–10092, 1993.
- [31] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," *The Journal of Chemical Physics*, vol. 103, no. 19, pp. 8577–8593, 1995.
- [32] S. Nosé, "A molecular dynamics method for simulations in the canonical ensemble," *Molecular Physics*, vol. 52, no. 2, pp. 255–268, 1984.
- [33] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," *Phys. Rev. A*, vol. 31, pp. 1695–1697, Mar 1985.
- [34] R. Clausius, "On a mechanical theorem applicable to heat," *Philosophical Magazine*, vol. 40, pp. 122–127, 1870.
- [35] J. Maxwell, "On reciprocal figures, frames and diagrams of forces," *Transactions of the Royal Society of Edinburg*, vol. XXVI, pp. 1–43, 1870.

## BIBLIOGRAPHY

- [36] M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids*. Clarendon Press, Oxford Science Publications, 1987.
- [37] H. Grubmüller, B. Heymann, and P. Tavan, “Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force,” *Science*, vol. 271, no. 5251, pp. 997–999, 1996.
- [38] M. Sotomayor and K. Schulten, “Single-Molecule Experiments in Vitro and in Silico,” *Science*, vol. 316, no. 5828, pp. 1144–1148, 2007.
- [39] W. Stacklies, F. Xia, and F. Gräter, “Dynamic Allostery in the Methionine Repressor Revealed by Force Distribution Analysis,” *PLoS Comput Biol*, vol. 5, p. e1000574, 11 2009.
- [40] W. Stacklies, M. C. Vega, M. Wilmanns, and F. Gräter, “Mechanical Network in Titin Immunoglobulin from Force Distribution Analysis,” *PLoS Comput Biol*, vol. 5, p. e1000306, 03 2009.
- [41] S. A. Edwards, J. Wagner, and F. Gräter, “Dynamic Prestress in a Globular Protein,” *PLoS Comput Biol*, vol. 8, p. e1002509, 05 2012.
- [42] W. Stacklies, C. Seifert, and F. Gräter, “Implementation of force distribution analysis for molecular dynamics simulations,” *BMC Bioinformatics*, vol. 12, no. 1, p. 101, 2011.
- [43] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, “GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation,” *Journal of Chemical Theory and Computation*, vol. 4, no. 3, pp. 435–447, 2008.
- [44] W. Humphrey, A. Dalke, and K. Schulten, “VMD – Visual Molecular Dynamics,” *Journal of Molecular Graphics*, vol. 14, pp. 33–38, 1996.
- [45] B. I. Costescu and F. Gräter, “Time-Resolved Force Distribution Analysis,” *BMC Biophysics*, 2013. In press.
- [46] S. Izvekov and G. A. Voth, “A Multiscale Coarse-Graining Method for Biomolecular Systems,” *The Journal of Physical Chemistry B*, vol. 109, no. 7, pp. 2469–2473, 2005.
- [47] A. P. Thompson, S. J. Plimpton, and W. Mattson, “General formulation of pressure and stress tensor for arbitrary many-body interaction potentials under periodic boundary conditions,” *The Journal of Chemical Physics*, vol. 131, no. 15, p. 154107, 2009.

- [48] O. H. S. Ollila, H. J. Risselada, M. Louhivuori, E. Lindahl, I. Vattulainen, and S. J. Marrink, "3D Pressure Field in Lipid Membranes and Membrane-Protein Complexes," *Phys. Rev. Lett.*, vol. 102, p. 078101, Feb 2009.
- [49] H. P. Boehm, A. Clauss, G. O. Fischer, and U. Hofmann, "Das Adsorptionsverhalten sehr dünner Kohlenstoff-Folien," *Zeitschrift für anorganische und allgemeine Chemie*, vol. 316, no. 3-4, pp. 119–127, 1962.
- [50] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, "Electric Field Effect in Atomically Thin Carbon Films," *Science*, vol. 306, no. 5696, pp. 666–669, 2004.
- [51] A. K. Geim, "Graphene: Status and Prospects," *Science*, vol. 324, no. 5934, pp. 1530–1534, 2009.
- [52] F. Chen and N. J. Tao, "Electron Transport in Single Molecules: From Benzene to Graphene," *Accounts of Chemical Research*, vol. 42, no. 3, pp. 429–438, 2009.
- [53] L. A. Ponomarenko, F. Schedin, M. I. Katsnelson, R. Yang, E. W. Hill, K. S. Novoselov, and A. K. Geim, "Chaotic Dirac Billiard in Graphene Quantum Dots," *Science*, vol. 320, no. 5874, pp. 356–358, 2008.
- [54] J. Tersoff, "Energies of fullerenes," *Phys. Rev. B*, vol. 46, pp. 15546–15549, Dec 1992.
- [55] D. W. Brenner, "Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films," *Phys. Rev. B*, vol. 42, pp. 9458–9471, Nov 1990.
- [56] D. W. Brenner, O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni, and S. B. Sinnott, "A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons," *Journal of Physics: Condensed Matter*, vol. 14, no. 4, p. 783, 2002.
- [57] S. Stuart, A. Tutein, and J. Harrison, "A reactive potential for hydrocarbons with intermolecular interactions," *Journal of Chemical Physics*, vol. 112, pp. 6472–6486, APR 8 2000.
- [58] J. H. Los, L. M. Ghiringhelli, E. J. Meijer, and A. Fasolino, "Improved long-range reactive bond-order potential for carbon. I. Construction," *Phys. Rev. B*, vol. 72, p. 214102, Dec 2005.
- [59] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, "ReaxFF: A Reactive Force Field for Hydrocarbons," *The Journal of Physical Chemistry A*, vol. 105, no. 41, pp. 9396–9409, 2001.

## BIBLIOGRAPHY

- [60] K. D. Nielson, A. C. T. van Duin, J. Oxgaard, W.-Q. Deng, and W. A. Goddard, "Development of the ReaxFF Reactive Force Field for Describing Transition Metal Catalyzed Reactions, with Application to the Initial Stages of the Catalytic Formation of Carbon Nanotubes," *The Journal of Physical Chemistry A*, vol. 109, no. 3, pp. 493–499, 2005.
- [61] B. I. Costescu, I. B. Baldus, and F. Gräter, "First Principles Based Morse Potential for Efficient Molecular Dynamics Simulations of Graphene," 2013. Submitted.
- [62] W. Qin, X. Li, W.-W. Bian, X.-J. Fan, and J.-Y. Qi, "Density functional theory calculations and molecular dynamics simulations of the adsorption of biomolecules on graphene surfaces," *Biomaterials*, vol. 31, no. 5, pp. 1007–1016, 2010.
- [63] R. R. Johnson, A. T. C. Johnson, and M. L. Klein, "Probing the Structure of DNA-Carbon Nanotube Hybrids with Molecular Dynamics," *Nano Letters*, vol. 8, no. 1, pp. 69–75, 2008.
- [64] R. Parthasarathi, N. R. Tummala, and A. Striolo, "Embedded Single-Walled Carbon Nanotubes Locally Perturb DOPC Phospholipid Bilayers," *The Journal of Physical Chemistry B*, vol. 116, no. 42, pp. 12769–12782, 2012.
- [65] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, "The Amber biomolecular simulation programs," *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1668–1688, 2005.
- [66] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, "CHARMM: The biomolecular simulation program," *Journal of Computational Chemistry*, vol. 30, no. 10, pp. 1545–1614, 2009.
- [67] T. Belytschko, S. P. Xiao, G. C. Schatz, and R. S. Ruoff, "Atomistic simulations of nanotube fracture," *Phys. Rev. B*, vol. 65, p. 235430, Jun 2002.
- [68] F. Scarpa, S. Adhikari, A. J. Gil, and C. Remillat, "The bending of single layer graphene sheets: the lattice versus continuum approach," *Nanotechnology*, vol. 21, no. 12, p. 125702, 2010.

- [69] W. H. Duan and C. M. Wang, “Nonlinear bending and stretching of a circular graphene sheet under a central point load,” *Nanotechnology*, vol. 20, no. 7, p. 075702, 2009.
- [70] I. W. Frank, D. M. Tanenbaum, A. M. van der Zande, and P. L. McEuen, “Mechanical properties of suspended graphene sheets,” *Journal of Vacuum Science & Technology B*, vol. 25, no. 6, pp. 2558–2561, 2007.
- [71] R. Faccio, P. A. Denis, H. Pardo, C. Goyenola, and Á. W. Mombrú, “Mechanical properties of graphene nanoribbons,” *Journal of Physics: Condensed Matter*, vol. 21, no. 28, p. 285304, 2009.
- [72] C. D. Reddy, S. Rajendran, and K. M. Liew, “Equilibrium configuration and continuum elastic properties of finite sized graphene,” *Nanotechnology*, vol. 17, no. 3, p. 864, 2006.
- [73] D. van der Spoel, E. Lindahl, B. Hess, A. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. Feenstra, R. van Drunen, and H. J. C. Berendsen, *GROMACS User Manual version 4.5.4*. 2010.
- [74] V. B. Shenoy, C. D. Reddy, A. Ramasubramaniam, and Y. W. Zhang, “Edge-Stress-Induced Warping of Graphene Sheets and Nanoribbons,” *Phys. Rev. Lett.*, vol. 101, p. 245501, Dec 2008.
- [75] R. C. Thompson-Flagg, M. J. B. Moura, and M. Marder, “Rippling of graphene,” *EPL (Europhysics Letters)*, vol. 85, no. 4, p. 46002, 2009.
- [76] S. Plimpton, “Fast parallel Algorithms for Short-Range Molecular Dynamics,” *Journal of Computational Physics*, vol. 117, no. 1, pp. 1 – 19, 1995.
- [77] N. Abedpour, M. Neek-Amal, R. Asgari, F. Shahbazi, N. Nafari, and M. R. R. Tabar, “Roughness of undoped graphene and its short-range induced gauge field,” *Phys. Rev. B*, vol. 76, p. 195407, Nov 2007.
- [78] A. Fasolino, J. H. Los, and M. I. Katsnelson, “Intrinsic ripples in graphene,” *Nat Mater*, vol. 6, pp. 858–861, Nov. 2007.
- [79] F. Scarpa, S. Adhikari, and A. S. Phani, “Effective elastic mechanical properties of single layer graphene sheets,” *Nanotechnology*, vol. 20, no. 6, p. 065709, 2009.
- [80] R. Grantab, V. B. Shenoy, and R. S. Ruoff, “Anomalous Strength Characteristics of Tilt Grain Boundaries in Graphene,” *Science*, vol. 330, no. 6006, pp. 946–948, 2010.

## BIBLIOGRAPHY

- [81] R. Khare, S. L. Mielke, J. T. Paci, S. Zhang, R. Ballarini, G. C. Schatz, and T. Belytschko, “Coupled quantum mechanical/molecular mechanical modeling of the fracture of defective carbon nanotubes and graphene sheets,” *Phys. Rev. B*, vol. 75, p. 075412, Feb 2007.
- [82] A. Irbäck, S. Mitternacht, and S. Mohanty, “Dissecting the mechanical unfolding of ubiquitin,” *Proc Natl Acad Sci U S A*, vol. 102, pp. 13427–13432, Sep 2005.
- [83] M. Cieplak and P. E. Marszalek, “Mechanical unfolding of ubiquitin molecules,” *J Chem Phys*, vol. 123, p. 194903, Nov 2005.
- [84] H. Lannon, E. Vanden-Eijnden, and J. Brujić, “Force-clamp analysis techniques give highest rank to stretched exponential unfolding kinetics in ubiquitin,” *Biophys J*, vol. 103, pp. 2215–2222, Nov 2012.
- [85] M. Clusel and E. I. Corwin, “Unfolding proteins with an atomic force microscope: force-fluctuation-induced nonexponential kinetics,” *Phys Rev E Stat Nonlin Soft Matter Phys*, vol. 84, p. 041920, Oct 2011.
- [86] S. Nauli, B. Kuhlman, I. Le Trong, R. E. Stenkamp, D. Teller, and D. Baker, “Crystal structures and increased stabilization of the protein g variants with switched folding pathways nug1 and nug2,” *Protein Sci*, vol. 11, pp. 2924–2931, Dec 2002.
- [87] S. Vijay-Kumar, C. E. Bugg, and W. J. Cook, “Structure of ubiquitin refined at 1.8 a resolution,” *J Mol Biol*, vol. 194, pp. 531–544, Apr 1987.
- [88] C. M. Pickart, “Mechanisms underlying ubiquitination,” *Annu Rev Biochem*, vol. 70, pp. 503–533, 2001.
- [89] L. Hicke and R. Dunn, “Regulation of membrane protein transport by ubiquitin and ubiquitin-binding proteins,” *Annu Rev Cell Dev Biol*, vol. 19, pp. 141–172, 2003.
- [90] S. Nauli, B. Kuhlman, and D. Baker, “Computer-based redesign of a protein folding pathway,” *Nat Struct Mol Biol*, vol. 8, pp. 602–605, July 2001.
- [91] J. Li, J. M. Fernández, and B. J. Berne, “Water’s role in the force-induced unfolding of ubiquitin,” *Proc Natl Acad Sci U S A*, vol. 107, pp. 19284–19289, Nov 2010.
- [92] A. Das and C. Mukhopadhyay, “Mechanical unfolding pathway and origin of mechanical stability of proteins of ubiquitin family: an investigation by steered molecular dynamics simulation,” *Proteins*, vol. 75, pp. 1024–1034, Jun 2009.
- [93] Z. T. Yew, T. McLeish, and E. Paci, “New dynamical window onto the landscape for forced protein unfolding,” *Phys Rev Lett*, vol. 101, p. 248104, Dec 2008.

- [94] S. Khorasanizadeh, I. D. Peters, T. R. Butt, and H. Roder, "Folding and Stability of a Tryptophan-Containing Mutant of Ubiquitin," *Biochemistry*, vol. 32, pp. 7054–7063, July 1993.
- [95] C. Debès, M. Wang, G. Caetano-Anollés, and F. Gräter, "Evolutionary optimization of protein folding," *PLoS Comput Biol*, vol. 9, p. e1002861, Jan 2013.
- [96] H. T. Young, S. A. Edwards, and F. Gräter, "How fast does a signal propagate through proteins ?," *PLoS One*, 2013.
- [97] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *Journal of Chemical Physics*, vol. 126, p. 014101, JAN 7 2007.
- [98] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, "The missing term in effective pair potentials," *The Journal of Physical Chemistry*, vol. 91, no. 24, pp. 6269–6271, 1987.
- [99] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *Journal of Applied Physics*, vol. 52, no. 12, pp. 7182–7190, 1981.
- [100] S. Nosé and M. Klein, "Constant pressure molecular dynamics for molecular systems," *Molecular Physics*, vol. 50, no. 5, pp. 1055–1076, 1983.
- [101] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259 – 268, 1992.
- [102] E. Jones, T. Oliphant, P. Peterson, *et al.*, "SciPy: Open source scientific tools for Python," 2001.
- [103] T. E. Oliphant, "Python for scientific computing," *Computing in Science and Engg.*, vol. 9, pp. 10–20, May 2007.
- [104] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science and Engg.*, vol. 9, pp. 90–95, May 2007.



---

# DECLARATION

---

I hereby declare that this thesis was authored entirely by me. Information derived from the published or unpublished work of others has been acknowledged in the text, and sources have been listed. This thesis has not been previously presented in identical or similar form to any other German or foreign examination board.

The work presented in this thesis was conducted between October 2009 and May 2013 at the Heidelberg Institute for Theoretical Studies gGmbH under the supervision of Dr. Frauke Gräter.

Heidelberg,

---

Ion Bogdan Costescu