Statistical Study of the Unfolding of Multimodular Proteins and their Energy Landscape by Atomic Force Microscopy

THÈSE N° 5440 (2012)

PRÉSENTÉE LE 21 SEPTEMBRE 2012
À LA FACULTÉ DES SCIENCES DE BASE
LABORATOIRE DE PHYSIQUE DE LA MATIÈRE VIVANTE
PROGRAMME DOCTORAL EN PHYSIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Fabrizio BENEDETTI

acceptée sur proposition du jury:

Prof. G. Meylan, président du jury
Prof. G. Dietler, Dr S. Sekatski, directeurs de thèse
Prof. P. De Los Rios, rapporteur
Prof. C. Micheletti, rapporteur
Prof. B. Samori, rapporteur

ii

Abstract

The aim of the present thesis is to investigate several aspects of: the proteins mechanics, interprotein interactions and to study also new techniques, theoretical and technical, to obtain and analyze the force spectroscopy experiments.

The first section is dedicated to the statistical properties of the unfolding forces in a chain of homomeric multimodular proteins. The basic idea of this kind of statistic is to divide the peaks observed in a force extension curve in separate groups and then analyze these groups considering their position in the force curves. In fact in a multimodular homomeric protein the unfolding force is related to the number of not yet unfolded modules (we call it "N"). Such effect yields to a linear dependence of the most probable unfolding force of a peak on ln(N). We demonstrate how such dependence can be used to extract the kinetic parameters and how, ignoring it, could lead to significant errors. Following this topic we continue with non kinetic methods that, using the resampling from the rupture forces of any peak, could reconstruct the rupture forces for all the other peaks in a chain.

Then a discussion about the Monte Carlo simulation for protein pulling is present. In fact a theoretical framework for such methodology has to be introduced to understand the various simulations done. In this chapter we also introduce a methodology to study the ligand receptor interactions when we directly functionalize the AFM tip and the substrate. In fact, in many of our experiments, we see a "cloud of points" in the force vs loading rate graph. We have modeled a system composed by "N" parallel springs, and studying the distribution of forces obtained in the force vs loading rate graph we have establish a procedure to restore the kinetic parameters used. Such procedure has then been used to discuss real experiments similar to biotin-avidin interaction.

In the following chapter we discuss a first order approximation of the Bell-Evans model where a more explicit form of the potential is considered. In particular the dependence of the curvature of the potential on the applied force at the minimum and at the metastable state is considered. In the well known Bell-Evans model the prefactors of the transition rate are fixed at any force, however this is not what happen in nature, where the prefactors (that are the second local derivative of the interacting energy with respect to the reaction coordinate in its minimum and maximum) depend on the force applied.

The results obtained with the force spectroscopy of the Laminin-binding-protein are discussed, in particular this protein showed a phase transition when

the pH was changed. The behavior of this protein changes, from a normal WLC behavior to a plateau behavior. The analysis of the force spectroscopy curves shows a distribution of length where the maximum of the first prominent peak correspond to the full length of the protein. However, length that could be associated with dimers and trymers are also present in this distribution.

Later a new approach to study the lock and key mechanism, using "handles" with a specific force extension pattern, is introduced. In particular handles of $(I27)_3$ and $(I27 - SNase)_3$ were biochemically attached to: strept-actin molecules, biotin molecules, RNase and Angiogenin. The main idea is to have a system composed by "handle-(molecule A)-(molecule B)-handle" where the handles are covalently attached to the respective molecules and the two molecules "A and B" are attached by secondary bonds. This approach allows a better recognition of the protein-protein interaction enabling us to filter out spurious events. Doing a statistic on the rupture forces and comparing this with the statistic of the detachments of the system of the bare handles, we are able to extract the informations of the interaction between the molecule A and B.

The two last chapters are of more preliminary character that the previous part of the thesis. A section is dedicated to the estimation of effective mass and viscous drag of the cantilevers studied by autocorrelation and noise power spectrum. Usually the noise power spectrum method is the most used, however the autocorrelation should give approximately the same information. The parameters obtained are important in high frequency modulation techniques. In fact, they are needed to interpret the results. The results of these two methods show a good agreement in the estimation of the mass and the viscous drag of the various cantilever used.

Afterwards a chapter is dedicated to the discussion of the force spectroscopy experiments using a low frequency modulation of the cantilever base. Such experiments allows us to record the phase and the amplitude shift of the modulation signal used. Using the amplitude channel we managed to restore the static force signal with a lower level of noise. Moreover these signals give us direct information about the dynamic stiffness and the lose of energy in the system, informations that, using the standard technique would be difficult (or even impossible) to obtain.

Abstract

Lo scopo della presente tesi é di investigare diversi aspetti della meccanica delle proteine e delle interazioni interproteina ed inoltre di studiare nuove tecniche, teoriche e tecniche, per ottenere ed analizzare gli esperimenti di spettroscopia di forza. La prima sezione é dedicata alle proprietá statistiche delle forze di unfolding in una catena di di proteine omomeriche. L'idea di base di questa statistica é quella di divivere i picchi osservati, in una curva di forza-estensione, in gruppi sepatati e di analizzare questi gruppi considerando la loro posizione nella curva di forza. In fatti in una proteina omomerica multimodulare la forza di unfolding é legata al numero di moduli non ancora unfold (chiamimano questo numero N). Questo effetto porta ad una dipendenza lineare della forza di rottura piú probabile per un picco con il ln(N). Dimostriamo quindi come tale dipendenza puó essere usata per estrarre i parametri cinetici e come, ignorandola, puó condurre ad errori significativi.

Seguendo questo tema continuiamo con un metodo non cinetico che, usando il ricampionamento delle forze di rottura da un qualsiasi picco, permette di ricostruire le forze di rottura per tutti gli altri picchi in una catena.

Successivamente una discussione sui metodi Monte Carlo per il tiraggio delle proteine é presentata. In fatti un quadro teorico per tale metodologia deve essere introdotto per capire meglio le varie simulazioni fatte. In questo capitolo introduciamo anche una metodologia per studiare le interazioni ligando-recettore quando si funzionalizza direttamente la punta dell'AFM e il substrato. In fatti, in molti nostri esperimenti, vediamo una "nuvola di punti" nel grafico forza vs loading rate. Abbiamo modellizzato un sistema composto da "N" molle in parallelo e, studiando la distribuzione delle forze ottenute nel grafico forza vs loading rate, abbiamo stabilito una procedura per riottenere i parametri cinetici usati. Questa procedura é quindi stata usata per discutere esperimenti veri simili a quelli di interazione di biotina-avidina.

Nel capitolo successivo discutiamo un'approssimazione al primo ordine del modello di Bell-Evans, dove una forma esplicita del potenziale viene considerata. In particolare la dipendenza della curvatura del potenziale al minimo e al punto metastabile é considerata. Nel modello di Bell-Evans i prefattori del tasso di transizione sono fissati rispetto alla forza, ma questo non é quello che succede in natura, dove i prefattori (che sono la derivata seconda del potenziale rispetto alla coordinata di reazione al minimo e al massimo) dipendono dalla forza applicata.

I risultati ottenuti con la spettroscopia di forza sulla Laminina (LBP) sono

presentati. In particolare questa proteina presenta una transizione di fase al cambio del pH. Il comportamento di questa proteina cambia, da un comportamento che segue la WLC ad un comportamento di tipo plateau. L'analisi delle curve di forza mostra una distribuzione di lunghezza dove il massimo del picco piú prominente corrisponde alla lunghezza totale della proteina. Sono anche presenti ulteriori picchi associabili a dimeri e trimeri.

Successivamente un nuovo approcio per studiare i meccanismi ligando-recettore, usando delle "maniglie" con uno specifico pattern nella curva di forza, é introdotto. In particolare queste "maniglie" proteiche di $(I27)_3$ e $(I27 - SNase)_3$ sono state biochimicamente attaccate a: molecole di strept-attina, molecole di biotina, RNase e Angiogenina.

L'idea é di avere dei sistemi composti da "maniglia-(molecola A)-(molecola B)-maniglia", dove le maniglie sono covalentemente attaccate alle rispettive molecole e dove le molecole "A e B" sono attaccate da legami secondari.

Questo approcio ci permette un migliore riconoscimento delle interazioni proteina-proteina permettendoci di filtrare gli eventi spuri. Facendo la statistica delle forze di rottura e comparando questa con la statistica delle forze di rotture dei sistemi composti da "maniglie" senza altre molecole, siamo capaci di estrarre informazione sulle interazioni tra le molecole A e B.

Gli ultimi due capitoli sono piú preliminari dei precedenti. Una sezione é dedicata alla stima della massa effettiva e del parametro di attrito viscoso dei cantilevers studiati con l'autocorrelazione e la densitá spettrale del rumore. Generalmente la densitá spettrale del rumore é il metodo preferenziale, ma l'autocorrelazione dovrebbe dare, approssimativamente, le stesse informazioni. I parametri ottenuti sono importanti negli sperimenti di modulazione ad alta frequenza. In fatti sono necessari per interpretare i risultati. I risultati ottenuti, tramite densitá spettrale e autocorrelazione, mostrano un buon accordo nella stima di massa e attrito viscoso dei vari cantilever utilizzati.

In fine un capitolo é dedicato alla discussione degli esperimenti di spettroscopia di forza utilizzando una modulazione a bassa frequenza della base del cantilever. Questi esperimenti ci permettoni di registrare i cambiamenti di fase e ampiezza della modulazione usata. Usando il canale di ampiezza siamo riusciti a ristorare il segnale di forza con un rumore piú basso. Inoltre questi segnali ci danno direttamente informazioni sull'elasticit dinamica e sulle perdite di energia del sistema, informazioni che, usando le tecniche standard sarebbero difficili (o addirittura impossibili) da ottenere.

Parole chiave: Miscroscopio a forza atomica (AFM), spettroscopia di forza a singola molecola, velocity clamp, simulazioni Monte Carlo, modellizazione barriera di energia, metodi non cinetici per spettroscopia di forza.

Acknowledgments

I remember when I come in Switzerland the first time in 2008. From Bologna I took the night train and I passed the border just with my driving license as document when they asked me the passport. The night was horrible, I spent it in a couchette in a vagon full of old and snoring guys. When I "woke" up at least I had a good surprise. The marvelous Switzerland landscape was out there, lakes, mountains, small waterfalls... At Lausanne I was really lucky to find the path to the EPFL, I was almost lost but still I managed to arrive. Tense and "shaked" I presented to Giovanni. I did my presentation in front of his full group. The presentation last for fifteen twenty minutes when, usually it should last for an hour. The presentation was a mess, I did prepared a trace to read (a trace in a paper to read! yes!) and I was able to get confused. At the end I had to explain a bootstrap method to a person with a really strong russian accent. The problem was also my english that, at that time, was still not good.

Giovanni. Thanks for have hired me and to have give me the chance to do science. If I was in you I would have not did the same. And thanks for the good environment, the low stress, the freedom in the research and the means that you gave me.

My second thanks go, then, to the aforementioned russian person. Sergey Sekatskii, one of the best physicist that I have the honor to know. His preparation, mental flexibility and his good character made my work interesting and stimulating. Without his work I would have accomplished much much less. I hope to have a chance to continue to collaborate with him in the future.

My acknowledgments then go to the polish part of the lab, mainly to Andrzej Kulik but I have also to mention Malgorzata Lekka and Piotr Marszalek with his PhD students. Andrzej to have helped me in different task and have propose the low frequency modulation experiments. Piotr to have host me in his lab during 4 months in 2011 and for the articles where I appear as author, Malgorzata to have taught some techniques for the protein-protein interactions. A thanks go also to Piotr' student, Minkyu Kim, with which I have profitable collaborate.

I have to mention Andrea Cerreta and Dusan Vobornik, whose company was important. Thank you guys for the uncountable (roughtly 400) number of lunch that we did together. Your company, your suggestions and your help has been important during these years, it is nice to have somebody to speak with of almost any kind of topic. A big thanks to Christine Vuichoud, her useful and precise job made my life much easier. And thanks for helping me with my terrible french even if you were not oblige to. A thanks go to Cristian Micheletti. The collaboration with him brought me an article in BJ. He is kind of a career-model for me (I hope that he is not going to destroy me during the private defense). I really hope to continue to work with him. A thanks go also to his student R. Potestio, who wrote and sent me the notes for the stochastic oscillator. Also Giovanni Bussi should be mentioned, his collaboration was fundamental for the BJ article.

Michel Kessous have to be mentioned for his continuos support and help for my countless problems with the computers. His help and enthusiasm was

foundamental, it is thanks to him if I had the chance to run so many simulations.

A thanks go also to Gianni Longo, his case make me always remember how terrible is the situation in italy, where scientists does not have any assurance even after years and years of study and research.

My best wishes and thanks go to the new members of the lab, Andrey Mikhaylov and Simone Ruggeri. My thanks go then to the other members of the lab, some of them have already left the lab during my stay, in particular I should not forget: Kristian Rechendorff, Guillaume Witz, Munir Mensi and Kanat Dukenbayev.

And finally I have to thanks my family, my dad Ruggero, that several time come in Switzerland to help me when I was seriously ill, my mom Patrizia and my brother Giorgio. During these years their support and company has been fundamental. Unfortunately several time I had discussions with them, manly because of the stress and the lack of results. Thank you for your support and patience, I am doing what I have always said I want to do since I was four and everything it's thank to you.

Contents

List of Figures

List of symbols

AFM, Atomic Force Microscope,

GB1, streptococcal B1 immunoglobulin-binding domain of protein G,

k_B Boltzmann constant,

k_c Spring constant of the cantilever,

FJC, Freely Jointed Chain,

WLC, Worm Like Chain,

EFJC, Extended Freely Jointed Chain,

EWLC, extended Worm Like Chain,

l_K, Kuhn length,

l_p, Persistence Length,

l_c, Contour length,

k, stiffness of an element in the EFJC model.

Φ, Elastic parameter of the EWLC

λ, transition rate

ν, prefactor of the transition rate

$\omega_{min/max}$, local double derivative of the potential near the minima and the maxima.

k_{off}, natural off rate of a bond (it express the rate at which a bond naturally break),

k_{on}, natural on rate of bond (it express the rate at which a bond naturally form),

N_B, number of bounded molecules

N_D, number of dissociated molecules

$\dot{F}$, loading rate, often express also as $\dot{f}$

APTES, 3-Aminopropyl-triethoxysilane

TSG, Template Stripped GOld,

Δx, in the reaction coordinate it represent the distance between the minimum of potential of the bounded state and the metastable state

k_{eff}, Effective spring constant, it is the composition of the spring constant of the cantilever and of the molecule pulled

k_{mol}, Spring constant of the entire molecule,

k_{module}, Spring constant of a module in a chain of omomeric objects

k_{unf}, Spring constant of an unfolded module

MC-HL, Monte Carlo High Loading rate

MC-LL, Monte Carlo Low Loading rate

MC-DY Monte Carlo Dynamic, stiffnesis changed dynamically

BC, Back-calculation

γ, viscous drag

x_T position of the metastable state in the Langevin energy landscape

x_F position of the stable state in the Langevin energy landscape

ΔG_{FT} difference of energy between the stable state and the metastable state

ΔG_{FU} difference of energy between the unstable state and the stable state

l_c contour length

$f_{WLC}(vt)$ a force that follow the worm like chain function

$\bar{K}$ spring constant of the completely unfolded construct

$\langle f \rangle_n$ average rupture force for a peak in position "n"

β_c dissipation constant of the cantilever

β_i dissipation constant of the molecule

k_i spring constant of the molecule

"z" position of the cantilever base

"d" position of the tip relative to the substrate, it is also the end-to-end distance

LPB, Laminin Binding Protein

Chapter 1

Introduction

The work that I am going to present here is related to the properties of the biopolymers. The general idea of my project, and in general of force spectroscopy, applied in polymers pulling and in ligand receptor characterization, is simple and can be formulated with a question, how does a string behave under tensional stress? We could imagine this string, attached at the very end to a surface. Moreover, the string has the property to create bonds between one or more section of itself. To answer to the initial question we can pull this ideal string with a spring and measure the force vs the extension.

If we do such an experiment many times, with similar strings, we could infer the average properties of these strings, like: the strength of the various bonds, the stiffness of the strings, their average lengths and more, depending on what we want to study.

However to do such an experiment at the nanometer scale we need an appropriate instrument. The Atomic Force Microscope, also called AFM, was developed around 1980 by Gerd Binnig, Calvin F. Quate and Heinrich Rohrer at the IBM Research in Zurich. The AFM is an instrument that allows to touch a sample with a small probe with nanometer precision. The probe, depending on the experiment that we want to perform, is in general composed by a cantilever with a small tip at the very end.

This cantilever is elastic and has a corresponding elastic constant. To explain its behavior we can make a quick parallel with a spring attacched to a string. If we put under stress this latter string also the spring will elongate allowing us to measure the force applied to the system. The cantilever, similarly to a spring, if attached to a string and if put under stress, will change its position. From this we can measure the force applied.

The probe is then used to touch the sample with controlled displacements.

The displacements, in the instrument, are controlled by piezoelectric elements that allow to achieve nanometer precision.

The upper surface of the cantilever is reflecting. The bending of the cantilever is measured by focusing a laser on the top of such cantilever. The laser beam will change trajectory as a function of the bending.

1
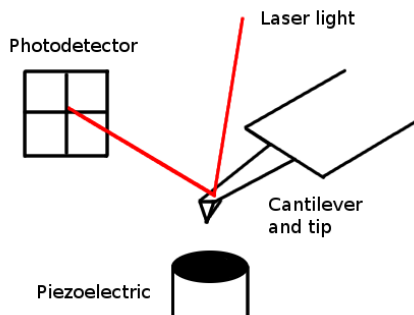
Figure 1.1: Diagram of an AFM.

An idealized scheme of a typical AFM can be seen in Fig.1.1.

This instrument is also extremely versatile, it allows to pull single molecules but it has been used for many other applications. In fact the original idea (that has been successfully applied since its invention) was to image samples (at the nanometer scale) by "touching" (e.g. [1, 2, 3]). However this argument goes beyond the scope of this thesis.

1.1 The study of the biopolymers

The biophysicists apply the quantitative and scientific methods to living matter. Of course, as in any other field, we often use simplifications. This approach allows us to obtain fundamental informations about the system that we are studying.

The main objective of this thesis is to study the inter- and intra-molecular forces involved, respectively, into the formation of complexes and in the structure of the biopolymers involved.

Moreover we did not only study the properties and the forces in the biopolymers but we have also tried to improve the experimental techinique for the obtaining of the data.

First we have to note that the biopolymers are omnipresent in each living organism (we can also include viruses here): DNA, protein, polysaccharides all have a role (often also more than one). It has been recognized that 70% of the proteins (and recently RNA) carry a structure that is deeply related to the sequence [4][5] (of course unstructured proteins carry a biological function, however the structure-function relationship of these proteins is more involved). For small globular proteins this is also know as "Anfinsen's dogma" a postulate first made by Christian B. Anfinsen, for which experimental demonstration he obained the Nobel Prize. The "dogma" says that, in physiological conditions (the conditions in which that protein is supposed to work) the native structure is determined by the amino acid sequence of the protein itself. This structure is unique, it is stable and it correspond to a kinetically accessible minimum of

α-helix   β-sheet
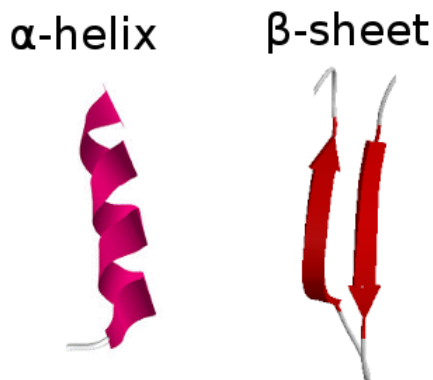
Figure 1.2: The two secondary structures.

the energy landscape.

Also if the structure is unique there are a number of common feature between the proteins. In fact, recurring structural motifs can be observed in different proteins.

The sequence of the aminoacids is simply called primary structure. Particular sequences of aminoacids give rise to a local structure that is called "secondary structure". These structures are three-dimensional motifs stabilized by hydrogen bonds.

The $\alpha-$helix is an helical conformation where, every "NH" group of each aminoacid at position n creates a hydrogen bond with the "CO" group of the residue $n-4$. The $\beta-$sheet is instead made by strands connected laterally by different hydrogen bonds. The various strands are made by stretched polypeptide chains, usually from three to ten aminoacids long.

The problem of the determination of the secondary structure, starting from the amminoacids sequence, is solved with an accuracy of 80% [6] using neural networks.

An example of the two secondary structures can be seen in Fig. 1.2.

The following level of organization is called tertiary structure and it is the spatial organization of the secondary motifs. Such organization is driven by different forces like hydrogen bonding, Van der Waals forces, electrostatics, hydrophobic/hydrophilic interactions (a protein in water tends to have an hydrophobic core, but proteins that are embedded in the hydrophobic cell membrane can behave differently).

Several methods exist to attempt predictions of the so called tertiary structure, in particular we have: *ab-initio*, by homology, side chain energy minimization, human pattern recognition ([7]), however, even if the landscape is dotted with successes, we are still unable to predict the native conformation of all the proteins.

The importance of the protein structure prediction resides in the fact that, if

we have the structure, we can better understand the function (and then possibly design a drug that "hits" a particular protein).

There is also another level of organization of the proteins that is called quaternary structure. In this organization a protein can interact with other proteins (also of the same type) to form a complex. However we should consider also the fact that RNA and protein can form complexes. An impressive RNA-protein complex is the ribosome (for which structural determination Venkatraman Ramakrishnan, Thomas A. Steitz and Ada E. Yonath got the Nobel Prize in 2009).

The function of most proteins is strictly related to the structure: it is the structure that regulates the chemical activity between a protein and a substrate. This structure recognition resembles, to a first approximation, the fit of key within a lock, where the mechanism works only if the contour of one fits the surface of its correspondent.

This process of recognition holds also for ligand-receptor interactions, and this process can be studied too with the AFM technique. The interactions within a ligand-receptor pair can be characterized by several parameters and can also have different pathways of formation.

The process of recognition shares a basic thermodynamic principle with the folding of a protein. In both cases the structures involved look for an energy minimum. This allows us to model both systems, again to a first approximation, with a simple two state model. Moreover, using the same model means also that we can use the same mathematics and obtain analogous parameters.

Many problems are related to the structure, or to the binding and the ligand-receptor interaction. For instance it is known that in pathogenic conditions a protein can misfold and form aggregates that may be toxic or pathogenic.

Understanding the conditions in which such processes occur, the kinetics and dynamics of the complex formation turn out to be a challenging task.

The formation of aggregates is just one of many examples: another is, for instance, the process of binding of virus to the cell membrane. As it is possible to imagine, this process is based on a lock-key mechanism. For example in the influenza virus, the virus uses a protein (the hemoagglutinin) to bind particular moiety on the cell surface (the lactosamyne). This is the first step of the infection mechanism and its characterization is important to study the virus and to design protection mechanisms.

The advantage of the technique that we are using is that it is possible to follow a single molecule at time. This allows us to have the information about the populations of conformers and the geometry of the molecules. Moreover we can also measure the kinetics of individual processes.

1.2 Possible applications

Why study the materials at the nanoscale?

One of the answers to this question comes directly from the mechanics of the muscular fibers. If we look at the composition of the vertebrate muscles we can see that titin is one of most abundant proteins. This protein plays a

role in the mechanical response. Some years ago it has been proposed that modules in titin could unfold reversibly in vivo to provide a mechanism for extension and entropical tensional resistance. Moreover the observations made with the AFM in velocity clamp and force clamp mode demonstrated that I27 (and similar modules) can refold in few seconds at zero or relatively low forces [8][9]. Minajeva et Al. [10] have shown that only few titin domains unfold at a time if force is applied. We can see that some ideas are arising to explain these observation, but the mechanics of this protein has yet to be fully understood.

Moreover we know that the evolution works by random mutations and natural selection, and therefore there are a number of open questions related to the evolution of titin in the muscles. Why natural selection has chosen this structure for the titin? Which advantages does it give? Why is titin composed by a sequence of similar, but not identical, protein modules with different mechanical resistances? To answer to these and other questions we need to study the properties of the components at the nanometer scale.

One can also try to look at the possible applications of the knowledge that came out from force spectroscopy. Recently it has been tried to create a material at the macroscopical scale carrying the properties of its components at nanometer scale.

Lv et Al. [11] created a material made of multimodular GB1 and resilin. This material showed elastic properties that mimicked the properties of the components. However, probably due to the particular "net like" structure of the crosslinks, the material was unable to stretch as much as the individual GB1 protein. Nevertheless this approach has shown the potential of force spectroscopy in material design and the need of a better understanding of the nanoscale physics, that could drive us to the production of new materials starting from a molecular level.

Chapter 2

Details of the AFM and of the experiments

2.1 AFM details

An AFM is composed by several parts: a piezoelectric actuator, an optic system, a photodetector and a cell or a holder to keep an external chip. In the chip we can find the "sensor" that allows us to measure the forces of our interactions. This "sensor" is a small flexible cantilever (a triangular cantilever can be approximatively 200x100 μm when a rectangular can be 200x20 μm, however dimensions are highly variable) at the side of the chip. At the very end of the cantilever we have a tip which usually, for force spectroscopy measurements, is pyramidal. The sharpness is an important parameter of the tip, in fact we want the tip to touch a really small area of our sample to interact, ideally, with a single molecule at a time. The common radius of an AFM tip is 30 nm, however the exact shape and the aperture depends on the specification.

The general scheme of an AFM was showed in Fig.1.1. When we approach a surface with a cantilever, the interaction forces bend the cantilever. The common way to detect the bending is to look at the laser beam reflected by the cantilever. This beam is sent to a four quadrant photodetector [12] that amplifies the signal and converts it in to a voltage, allowing not only to measure the vertical displacement but also the lateral (torsional) bending. In the first stage of an experiment we need to correlate the deflection of the cantilever with the signal obtained at the detector.

This is done by approaching the cantilever to the surface with the piezo and correlating the piezoelectric displacement (that is assumed to be correct, and can be calibrated independently) with the change of voltage. Now, when we know how much the cantilever bends as a function of distance, we can calculate its spring constant.

This is in general done using the equipartition theorem [13] considering the cantilever as an oscillator with one degree of freedom:

$$\frac{1}{2}k_B T = \frac{1}{2}k_c \langle x^2 \rangle \tag{2.1}$$

Where k_c is the spring constant of the cantilever and $\langle x^2 \rangle$ the time average square thermal fluctuation of the cantilever.

To distinguish the different contributions to the fluctuation we use the power spectrum density (PSD) of the signal. Looking at the PSD we can skim the spurious contribution to the deflection and limit our analysis at the region in which we have the peak of the resonance frequency of the cantilever.

The Parseval theorem allows us to write eq. 2.2.

$$\int |x(t)|^2 dt = \int |\tilde{x}(\omega)|^2 d\omega \tag{2.2}$$

The major contribution to the power spectrum of the cantilever (if no external sources excite it with a different frequency) is due to the resonance frequency of the cantilever. This procedure can be used in air or in liquid. During this procedure there are also other corrections that have to be taken into account (see also [14]).

2.1.1   Force curves

The force distance curve is the signal that we obtain from an AFM during a force pulling experiment. This signal show the force measured in function of the displacement of the base of the cantilever.

In the velocity clamp technique it is important to know exactly how the piezo moves, also the interpretation of the curves is easier if the displacement is linear with the time. For this purpose the AFM exploits a so-called "Close-loop", where a sensor control the position and a feedback make it linear with the voltage applied.

In this way we can obtain a constant velocity during the approach of the surface and a constant velocity during the retraction. Often this is required because we want the same pulling speed in different regions of the curve.

The drawing of a force curve can be seen in Fig. 2.2.

There are several protocols to obtain a force spectroscopy curve. Usually the starting position of the cantilever is far from the surface, then we approach the surface applying a voltage to the piezo that increases with time until it reaches a setpoint, after that we decrease the voltage until we get the initial value. This procedure can be modified: we can use different speeds for approaching and retracting, we can also keep the cantilever in contact at a given force for a given period and then retract back from the surface. Another interesting strategy is to revert the ramp starting from a contact position then do an extension and finally return to the contact. The ramp could be also more exotic, in fact it is possible to modulate the ramp [15] to observe mechanisms that would be hidden in a standard pulling sequence.

The "non contact" part contains the events that we study. In general these events resemble an asymmetrical peak (or a sawtooth pattern made of multiple,

consecutive peaks if, for example, we are pulling multimodular proteins) however the obtained signal depends on the structure and on the precise nature of the interactions. For instance the DNA molecule shows a long and very flat plateau around 70 pN ([16]) and only at the very end there is again an increase of the force.

The force curves are usually "corrected" changing the coordinate from the cantilever base into the extension of the molecule. This correction is done in the following way: $E(t) = D(t) - (F(t)/k_c)$ where "E" is the extension, "D" the coordinate of the base of the cantilever, F(t) is the force at the time "t" and k_c the spring constant of the cantilever.

2.1.2 Force extension models

The Freeely-Jointed Chain (FJC) is the simplest model to describe the behavior of a polymer. This model considers a polymer composed by consecutive identical rigid segments with random (uniform distributed) orientation in space. Substantially it corresponds to a three dimensional random walk with a fixed step size. Another aspect of this model is that the excluded volume effects are neglected. The total length of an ideal polymer, in this kind of model, is given by $L = Nl$, where "N" is the number of modules and "l" the length of a module. This kind of polymer fluctuates in time and so, if we take the average distance between two ends averaged over the time or if we take an ensemble of polymers we get: $\langle \vec{R} \rangle = \Sigma_{i=1}^{N} \langle \vec{r_i} \rangle = 0$

Each segment of the chain can be seen as a vector $\vec{r_n}$ and all of them are independent one from anoter. This tells us that, according to the central limit theorem, $\vec{R}$ is normally distributed. So all the 3D component of "R" are normally distributed with a mean of "0" and a variance given by eq. 2.3

$$\sigma^2 = \langle R_x^2 \rangle - \langle R_x \rangle^2 = \langle R_x^2 \rangle - 0 \tag{2.3}$$

Due to the isotropy we have: $\langle R_x^2 \rangle = \langle R_y^2 \rangle = \langle R_z^2 \rangle = N\frac{l^2}{3}$, this is due to the fact that $\langle R^2 \rangle = Nl^2 = Ll$ and applying the Pythagoras theorem we can easily see that the average end to end distance is $\sqrt{Ll}$. The end to end distance is distributed as: $P(\vec{R}) = \left(\frac{3}{2\pi Nl^2}\right)^{3/2} e^{-\frac{3\vec{R}^2}{2Nl^2}}$.

As aforementioned, the stretching of a polymer can be represented by a force extension curve. During the stretching we have two concurrent forces that oppose to the stretching. At small forces, when the protein is near the natural conformation, the entropy dominates the chain. So, elongating the chain means a decrease in the number of possible configurations of it, and to do it we need to overcome the entropy of the chain. At any temperature this entropy effect is reflected as a force. In an ideal process we need to give the energy $T\Delta S$ to elongate the chain.

At higher forces the contribution of the molecular bonds becomes significant, so in such case we are changing the enthalpy of the system and we speak about enthalpic elasticity. High forces can change the internal arrangement of the

molecule, leading to a change of conformation of one or more elements that compose the chain.

The relation between the extension of an ideal polymer and an applied force was first obtained by Kuhn [17] and it is given in eq. 2.4.

$$x(F) = L \left[coth \left(\frac{Fl_K}{k_B T} \right) - \frac{k_B T}{Fl_K} \right] \tag{2.4}$$

Freely jointed chain describes quite well the systems where each segment is independent from the previous, like in polysaccharides, however there are other systems that are better described by a model where we have continuous elements correlated each other. The most important of them is the Worm Like Chain (WLC). For a certain string of length "l" we can parametrize the path as "s" $\in (0, l)$. We can then define the tangent "t" to the string at a point "s" as $t(s)$. In such kind of string the correlation between the tangents decay exponentially.

$$\langle \hat{t}(s) \cdot \hat{t}(s') \rangle = exp(-|s - s'|/l_p) \tag{2.5}$$

In this decay we have introduced the length "l_p" called "persistence length". The persistence length measure the distance at which the correlations are lost, however this parameter can be related to the bending stiffness ("noted as "B_s") and the temperature as $l_p = B_s/k_B T$.

The WLC model gives the eq. 2.6 for the force response:

$$F(x) = \frac{k_B T}{l_p} \left[\frac{1}{4(1 - \frac{x}{l_c})^2} + \frac{x}{l_c} - \frac{1}{4} \right] \tag{2.6}$$

In an article of Bustamante et al. [18] the authors show that, for λ-phage DNA, the FJC and the WLC coincide at low (< 0.1 pN) and high forces (> 50 pN), however for intermediate forces the WLC fit better the data.

The two model can be improved taking into account the module stiffness under force, this lead to the so called "Extensible Freely Jointed Chain" (EFJC) and the "Extensible Worm Like Chain" (EWLC), where respectively the therm "k" and "Φ" represent the stiffness of an element (more properly $\Phi \approx kL$).

The two equations for these models are given in eq. 2.7 and 2.8.

$$x(F) = L \left[coth \left(\frac{Fl_K}{k_B T} \right) - \frac{k_B T}{Fl_K} \right] \left(1 + \frac{F}{kL} \right) \tag{2.7}$$

$$F(x) = \frac{k_B T}{pl} \left[\frac{1}{4(1 - \frac{x}{l_c} + \frac{F}{\Phi})^2} + \frac{x}{l_c} - \frac{1}{4} - \frac{F}{\Phi} \right] \tag{2.8}$$

These two models turn to be useful to estimate the stiffness of the chain or of the modules, for instance an interesting approach to estimate the stiffness of the modules in a polysaccharide chain can be seen in [19].

2.2 The Bell-Evans model

In this thesis we use an interaction model that considers two states at different energy divided by a barrier at higher energy. This landscape describes the energy profile of the units of our biopolymers. We can model our biopolymers as composed by a series of two-states springs.

The model can be schematized as in fig. 2.4.

This model has been applied for chemistry systems for a long time and it has been first mathematically formalized by Kramers [20]. The model was refined by Bell[21] and then by Evans, who introduced a loading rate dependence of the lifetime of the bond in presence of a force. As Evans states in [22]: "*These bonds will fail under any level of pulling force if help for a sufficient time. Thus, when tested with ultrasensitive force probes, we expect cohesive material strength and strength of adhesion at interfaces to be time- and loading-rate dependent properties.*" .

In the following passages we will introduce some basics mathematics about the Bell-Evans theory. More details on how to derive the model can be found in the article of Kramer [20].

The first theoretical description of a transition or reaction rate was introduced by Arrhenius. Thanks to him we can write the dissociation rate λ as eq. 2.9.

$$\lambda = \nu exp \left(-\frac{\Delta G}{k_B T} \right) \tag{2.9}$$

This was later extended by Kramers, and ν become the natural vibration frequency of the bond determined by the frequency at the stable and metastable state divided by the damping (as written in eq.2.10). More precisely the two frequency are $\omega^2_{max/min} = |d^2 U(x)/dx^2|_{x=x_{max/min}}$.

$$\lambda = \frac{\omega_{min}\omega_{max}}{2\pi\gamma} exp \left(-\frac{\Delta G}{k_B T} \right) \tag{2.10}$$

When we apply a force the eq. 2.10 changes and we have:

$$\lambda(F) = \frac{\omega_{min}\omega_{max}}{2\pi\gamma} exp \left(-\frac{(\Delta G - F\Delta x)}{k_B T} \right) = k_{off} exp \left(\frac{F\Delta x}{k_B T} \right) \tag{2.11}$$

In eq. 2.11 we have grouped the vibration rates with the energy term to the k_{off} leaving the exponential term that depends on $F\Delta x$. Using such notation however change the meaning of the formula, passing from the Kramer theory, where we explicitly express the dependence of ω on the force, to the Bell-Evans theory, where this dependence is neglected.

Using the rate of dissociation we can write the number of molecules that dissociate at a time "t" (eq. 2.12).

$$\frac{dN_B}{dt} = -\lambda(t)N_B \tag{2.12}$$

The eq. 2.12 gives eq. 2.13

$$N_B(t) = exp\left(- \int_0^t dt' \lambda(t') \right) \tag{2.13}$$

Correspondingly we can calculate the number of dissociated bonds with eq. 2.14.

$$N_D(t) = 1 - N_B(t) = 1 - exp\left(- \int_0^t dt' \lambda(t') \right) \tag{2.14}$$

This latter equation can be written as function of the loading rate if we assume that the force is a continuous and monotonous function of the time, thus we have eq. 2.15

$$N_D(f) = 1 - exp\left(-\frac{1}{\dot{f}} \int_0^f df' \lambda(f') \right) \tag{2.15}$$

The probability distribution of bond breakages as a function of the force is given by the derivative of eq. 2.15 and (with some substitutions) give eq. 2.16.

$$p(F) = \lambda(F)s(F) = \frac{k_{off}}{\dot{f}} \exp\left(\frac{\Delta x F}{k_B T} \right) \exp\left[\frac{k_{off} k_B T}{\dot{f} \Delta x} \left(1 - \exp\left(\frac{\Delta x F}{k_B T} \right) \right) \right]$$

Deriving the probability distribution and putting it to zero we can find the most probable force as a function of the loading rate $\dot{F}$ as written in eq. 2.17. This latter equation has been used in many different papers to get the kinetic parameters of a protein or a complex using the dependence on the loading rate.

$$\frac{d}{df}\left(\frac{dP_D}{df}(F) \right) = 0 \tag{2.16}$$

$$F(\dot{F}) = \frac{k_B T}{\Delta x} ln\left(\frac{\Delta x \dot{F}}{k_{off} k_B T} \right) \tag{2.17}$$

The equation 2.17 correlates the rupture/unfolding forces with the loading rate and the kinetic parameters of a particular interaction. This correlation can be exploited to obtain the kinetic parameters changing the speed at which a bond is pulled (as example: [23],[24],[25],[26]).

The model that we have introduced is really useful for our work, however it is an extreme simplification and many other factors have to be taken into account to estimate the kinetic parameters. For instance we have investigated how the multiplicity of modules in a multimodular protein change the rate, also other effects like the polymer length and the temperature may play an important role. Moreover the vibrational terms and the position of the barrier change when a force is applied. Such effects have been investigated and results are reported in the following chapters.

2.2.1 Functionalization procedure to study ligand-receptor interactions

The procedure to functionalize the surface and the tip of a AFM cantilever depends mostly on the type of experiment that we want to perform.

The basic idea for protein-protein (or ligand-receptor) interaction AFM studies is to have one or more ligands on the AFM tip and the corresponding receptor on a flat surface. Both molecules have to be strongly attached to their own substrate, this is because the bond that keeps the proteins attached to the substrate does not have to compete with the ligand-receptor interaction.

The aim is to have a clear statistics based on only one kind of interaction.

To do so we use a protocol for surface functionalization. The first step is to use a silanization process on the mica surface (mica is chosen because it is easy to get atomically flat surfaces apt to SMFS experiments). The process we used is based on 3-Aminopropyl-triethoxysilane (APTES) and glutaraldehyde. APTES allows us to have amino groups on the surface of the mica, while the glutaraldehyde has two highly reactive aldehyde groups that react with NH_2 groups. The idea is in fact to lay a carpet of amino groups on the surface, then add glutaraldehyde that binds, with one of the two groups, the NH_2 on the surface. After this, adding a protein solution, the remaining aldehyde group binds the NH_2 exposed on the protein surface. In this way we have covalent bonds between the molecules and the surface.

It is important to notice that APTES polymerizes in presence of water. It is possible to use the APTES using also a wet strategy (a solution 0.1% v/v of APTES in water) but, for pulling, a good precaution is to use a solution freshly prepared, to avoid APTES polymers that would interfere with the experiment. With the wet strategy we deposit the aforementioned solution on mica for one minute and then we can wash with ultrapure water.

Another strategy is to use a vapour method, which consists in putting under vacuum: a Petri dish with the mica surface that we want to functionalize and in another Petri dish some droplet of APTES. Under vacuum, the APTES vapour saturates the environment and reacts with the mica surface, functionalizing it with amino groups. This strategy is more time consuming: experience shows that two hours are required for a good functionalization.

After the silanization we add a solution of glutaraldehyde (1% v/v) for 15 minutes. The surface is then rinsed with deionized water and then a solution of protein is put on the surface for 15 minutes. The surface is finally rinsed with PBS (phosphate buffer saline, pH 7.4, P4417 Sigma) to remove the unreacted protein. Following this procedure we have a sample ready to be used for the experiments, however we have to stress that the solution on the surface (after the deposition of the protein) should not be allowed to dry, because it would denature the proteins. In general also the protein solution should be treated gently, avoiding contact with air and violent shaking (a gentle shake can be performed with a pipette).

The same strategy could be used for the cantilever tip, where we use the corresponding protein for the interaction that we want to study.

2.2.2 Physisorption on gold surfaces

Physical absorption (physisorption) is a mechanism that can be exploited to attach proteins and polysaccharides on a substrate. This kind of interaction relies on secondary bonds (e.g. van der Waals) and it is unspecific. We have to precise that the exact nature of the bond that keep the molecules attached to the surface is unknown. In this case, we assume that the bonds are non-covalent because we did not use any other substance than the molecules of interest. For this reason we speak about physisorption,however, a small chance to create covalent bond is still present. In some of our experiments, we have absorbed proteins on gold surfaces. These gold surfaces, called Template Stripped Gold (TSG), are composed by a thin film of gold obtained by evaporating the metal on clean mica surface. Next the upper part of this layer is glued to a glass coverslip. In this way we have a flat face attached to the mica and the bottom face attached to a coverslip. With a simple mechanical operation it is possible to remove the mica from the TSG leaving the flat layer of gold on the glass. Physisorption is obtained simply by depositing a solution of protein on the surface. The quantity of protein adsorbed depends on the concentration and to the deposition time, however this process is highly stochastic and it cannot be easily predicted.

2.2.3 Plasma functionalization procedure

To obtain force distance curves with dextran and with multimodular proteins we found a really inexpensive method. Based on empirical observation we saw that a glass coverslip treated with a Bunsen flame acquires an hydrophilic character. We tried to use this procedure in some experiments, depositing a solution of multimodular protein on a treated coverslip and comparing the results obtained with untreated glass coverslip. The initial results were promising under the point of view of binding efficiency, but they were unfortunately "dirty" regarding the molecule pattern. We then tried to use a plasma cleaner for the glass coverslip, however the plasma may introduces residues and particles that create problems to the proteins deposited. After some attempts we found the best conditions to treat the coverslip. We washed the coverslips with a strong flux of water. Then we dried them with compress air and then they were treated using (Harrik Plasma cleaner PDC-32G) in a plasma of air at a pressure of 1000 ± 100 mTorr at low-medium intensity for 5 minutes. We notice that proteins and polysaccharides like dextran were able to stick on the surface giving also a clean, good SMFS signal.

A possible explanation for this behavior is that the plasma removes almost all the substances that were previously adsorbed on the surface. When we deposit the substance of interest we have a physical absorption but now, because the surface is really clean, we have many active sites that attract the substance of interest.

For long expositions (> 1 min) also the roughness of the surface can change, this factor can contribute too to the probability of absorption of a substance.

(Despite its utility we did not have time to study all the detail of the plasma functionalization. It is the opinion of the author that for biomolecular interaction study this topic would deserve more attention.)

For the dextran, and in general for the polysaccharides that can be dissolved in water, it is sufficient to deposit around 50 μl of solution at 5% v/v on a freshly treated coverslip and leave it dry overnight. The solution forms a thick and visible layer that has to be removed. Such layer is not usable for the force measurements, in fact, we want to pull a single molecule at time, but this would be impossible with the aforementioned layer.

Therefore the coverslip is extensively washed with a strong flux of ultrapure water (approx 200 ml) to remove the dessicated layer of polysaccharides from the surface. The sample is then dried again and it is ready for an experiment.

The procedure to prepare a sample with multimodular protein is similar. In this case we deposit 30-40 μl of a solution with a concentration of about 30 $\mu g/ml$ of protein for 20 minutes. The deposition is made inside a closed Petri dish where we also have put some droplet of water at a side of it, this is to create a humid environment and prevent the drying of the protein solution deposited.

After 20 minutes we can wash the unreacted protein adding and removing for two or three times 50 μl of the buffer solution that we use in the experiment (usually PBS). Here, in any case, we should prevent the drying of the surface. This is again because dry proteins exposed to air denature.
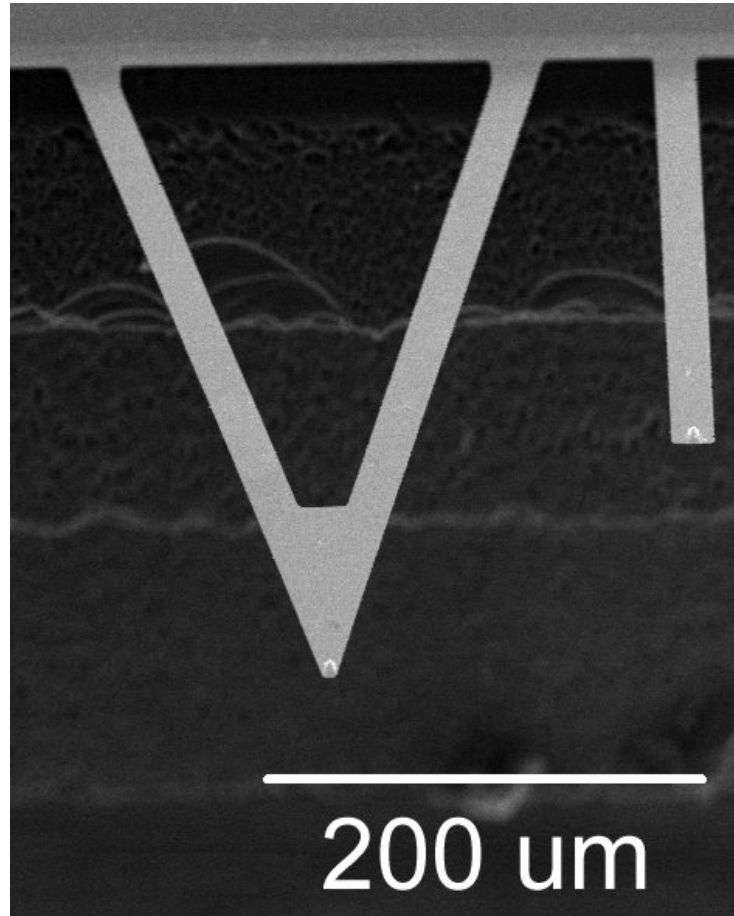
Figure 2.1: We can see two cantilevers of different geometry. At the very end of each one is possible to notice the tip. The different geometries correspond to different correction factors when determining the spring constant of the cantilever.
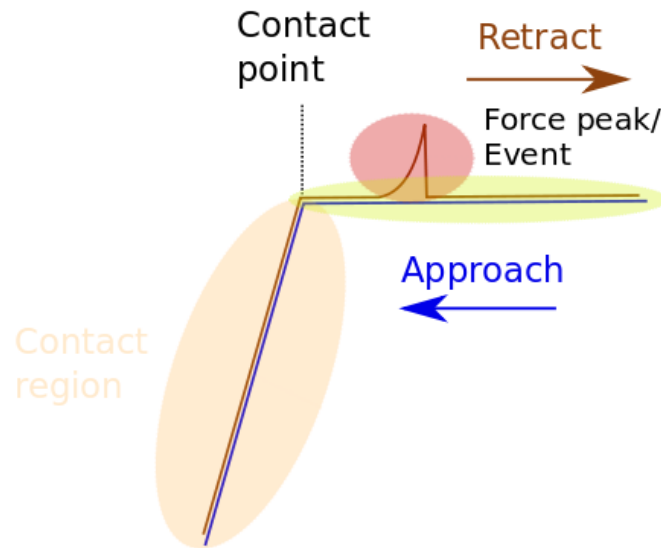
Figure 2.2: In this figure we can see an ideal force curve. At the beginning (blue curve) we have the approach, when the piezo moves the chip close to the sample. Until we are in the light yellow region we are "out of contact": the cantilever does not interact with the surface (just the fluid and the long range interaction if in air). At the contact point the cantilever touches the sample/surface and it starts to bend, increasing linearly the force signal with the displacement. At the end of the blue curve the retractions, now we are pulling away the chip from the surface. If the cantilever has "fished" something we will see the force event in the out of contact region. In this case we have depicted only a simple, unique peak event, however the type of events that can happen are diverse and they depend on the nature of the interaction.
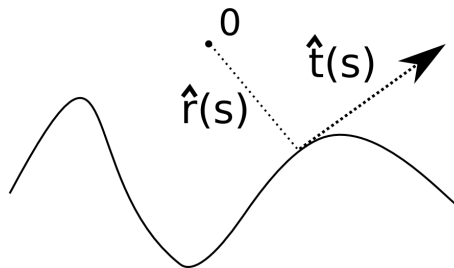
Figure 2.3
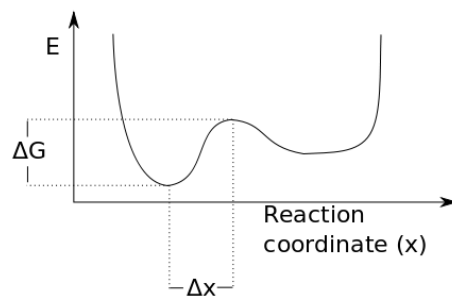
Figure 2.4: Here we can see an hypotetical potential that takes into account two states only: a bound state and an unbound state. In normal conditions, like in the common velocity clamp experiments, we cannot determine the parameters of the unbound state.
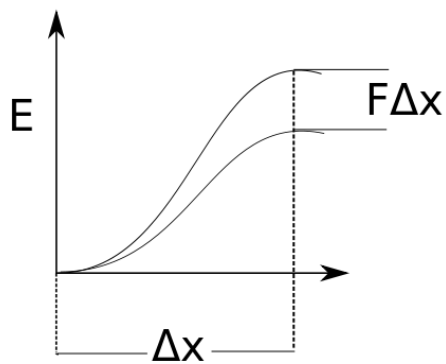
Figure 2.5: The energy barrier in the Evans model. The energy barrier is decreased by a factor $F\Delta x$ when a force is applied. We also notice that the x position of the maximum does not change.

Chapter 3

Multimodular proteins mechanics

In this chapter we will focus on the behavior under pulling of multimodular homomeric proteins. We will analyze their behavior by two different strategies. The first one is theoretical, where we will explain the effect of the domain multiplicity on the force statistics of the multimodular proteins, in particular we will focus on how to get the kinetic parameters by exploiting this effect. The second strategy is instead phenomenological and it will show how, under some assumptions, it is possible to predict the average rupture force for each module in a chain of "N" modules based on the knowledge of the properties of just one of them.

The two methods have been applied to two different data sets, nevertheless they give consistent results.

The methods have been applied to a multimodular construct of eight sequentially linked GB1 protein domains, that is, the streptococcal B1 immunoglobulin-binding domain of protein G [27].

3.1   A kinetic point of view

The section is dedicated to the methodology that we have developed to obtain the kinetic parameters of a protein module that builds a multimodular (homomeric) construct.

The main reason for it was the recognition of the fact that the probability to unfold of a module, independently, is the same for all the modules. However when modules are linked in a chain we have to take into account the fact that, at the same time, multiple modules withstand force -and therefore are subject to a non-negligible probability to unfold. This leads to an hierarchical behavior where the most probable unfolding force increases from the unfolding peak closer to the contact point to the last peak, just before the detachment.

Moreover, in the literature it is possible to find several statements about the difference in the forces of the peaks in a multimodular chain, however few have ever taken into account how this variable can be used to obtain information about the kinetic of the process.

Most published work in the field either used some global statistics of the unfolding peaks in the force-extension curves deduced from a linear regression method [28] or estimated the kinetic parameters from the distribution of the forces [29]. Only recently have some studies focused on the hierarchical behavior in multimodular systems [30, 31, 32, 33], see also [34].

A similar method can be seen in a work by Brockwell [30]; but in our analysis, we do not need a characteristic length scale (Δx) for the molecule and we do not calculate the loading rate to model the protein at every peak (for which these authors used the elasticity function of the Worm-Like Chain).

The main idea is that, when multimodular proteins composed of a number of identical modules are unfolded in the course of a single-molecule force-spectroscopy experiment, both the probability distribution of the unfolding force and the most probable unfolding force depend on the number of not-yet-unfolded modules in the protein.

The unfolding process can be described in the framework of a generalized Bell-Evans model, based on certain minimal assumptions concerning the force loading conditions.

In particular, the equations that we propose for the unfolding force predict, for a multimodular protein containing N not-yet-unfolded modules, a linear dependence of the most probable unfolding force on ln(N). This conclusion has been confirmed in experiments with GB1 proteins, which verify in addition further consequences of the proposed model, including a modified probability distribution of the unfolding force.

In this work, multimodular proteins are stretched and the most probable force, F , at which the modules unfold, known as the unfolding or rupture force, is determined.

When the rate of unfolding as a function of the applied stretching force, f, can be expressed by a known monotonic function $\lambda(f)$ (see eq.2.11 in chapter 2). The probability of module survival (*i.e.* the probability that a module has not yet unfolded during force loading from $f = 0$ to $f = F$) is given by the expression

$$s(F) = \exp\left(-\int_0^F \lambda(f)\frac{df}{\dot{f}} \right) \tag{3.1}$$

Correspondingly, the probability that the module unfolds in the force interval $[F, F + dF]$ is given by eq.3.2.

$$dp = p(F)dF = \lambda(F)s(F)dF = \lambda(F)\exp\left(-\int_0^F \lambda(f)\frac{df}{\dot{f}} \right) dF \tag{3.2}$$

If the force loading rate $\dot{f}$ is constant, differentiation of Eq. (3.2) with respect to f gives immediately the most probable unfolding or rupture force as written

in 3.3.

$$\dot{f}\frac{d\lambda}{df} = \lambda^2(f) \tag{3.3}$$

As said in the chapter 1, the rate of unfolding $\lambda(f)$ is given by eq. 2.11.

However, when we deal with multimodular proteins the situation is more complex.
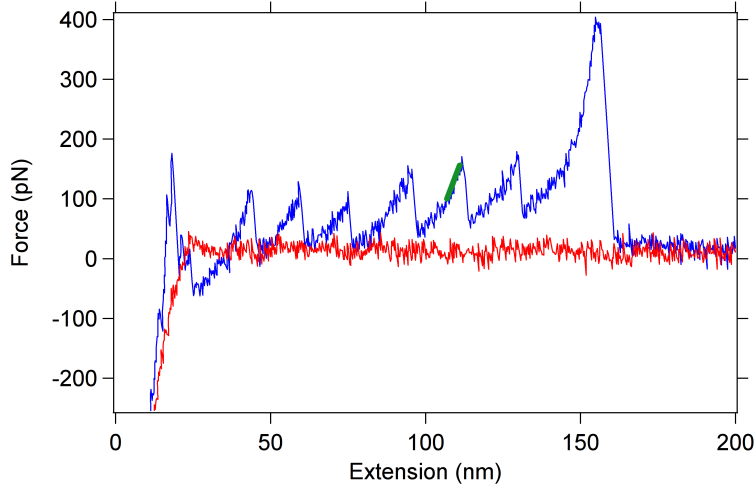
Figure 3.1: A typical force curve for velocity clamp $GB1_8$ stretching. In red is shown the approach curve and in blue the retraction curve with the unfolding peaks. The latter contains 6 unfolding events and the detachment event, where the molecule detaches from the AFM tip. In the hierarchy of forces, the unfolding force increases statistically as the number of intact modules on the molecule decreases. The dark-green line represents a linear fit near the unfolding peak, which is used to determine the approximate spring constant of the stretched molecule, and can be combined with the spring constant of the cantilever to calculate the effective loading rate.

Fig. 3.1 shows an example of typical single-molecule force-spectroscopy data for a multimodular protein. To deal with this type of system, we introduce the following generalization. In the presence of N identical modules connected in series, which are subject to the same stretching force F, we require instead of Eq. (3.2) an equation reflecting 1) that now not one but N modules must survive during force loading from $f = 0$ to $f = F$ (leading to a factor $s^N(F)$), and 2) that any one of these N modules (whence a further factor N) may unfold in the infinitesimal force interval $[F, F + dF]$. Thus

$$dp_N = p(N, F)dF = N\lambda(F)s(F)^N dF \tag{3.4}$$

differentiation of which yields, in this case, a most probable unfolding force of

$$F_{rupt}(N) = \frac{k_B T}{\Delta x} \ln \left(\frac{\dot{f}\Delta x}{N k_{off} k_B T} \right) \tag{3.5}$$

If now we suppose that, after one module has unfolded, (a) the force loading conditions are exactly the same for the next module, namely the force returns rapidly to zero and (b) force loading begins anew with the same constant rate $\dot{f}$ as before[1], we may apply Eq. (3.5) again, but now with $N-1$ modules. Thus we can compare the unfolding force $F_{rupt}(n)$ of the nth module with that of the mth module, and we will show that this comparison delivers essential new information about the system.

3.1.1   Materials and methods

Our experiments were performed on multimeric constructs consisting of eight GB1 modules [9, 35, 36] dissolved in Tris/HCl buffer (10 mM, pH 7.5) at a concentration of approximately 20 μg/ml and deposited on gold as described in "Introduction".

The lever used was an NP (Veeco) which, for the data set presented, was measured to have an elastic constant of 0.112 N/m. We have performed several series of experiments to study the unfolding of $(GB1)_8$ proteins under different pulling speed and with different cantilevers. To avoid possible uncertainties in the data, all of the experiments described here were recorded on the same sample under constant conditions and with a single cantilever. The construct was picked up by the tip at random positions. The constructs were pulled at a speed of $\dot{z} = 2180\ nm/s$. The curves obtained were analyzed using "Hooke"[37], an open-source software package designed to extract information from AFM curves. For our data analysis, we have selected only curves presenting a detaching peak, because otherwise it is impossible to determine the total number of domains picked up between the surface and the tip. In the following, we indexed the last unfolded module in a stretching curve as peak 1, where we assume that only one intact module was present during the last extension of the molecule, while peak N in an unfolding sequence is an unfolding event on a complex with N intact modules. The unfolding forces for all peaks of each order (from 1 to N) were then treated together: this ordering ensures that we are merging data about unfolding events which are mutually consistent concerning the number of intact modules. The force loading rate for each group is the average of the slopes of the peaks belonging to that group combined with the spring constant of the cantilever,

$$\dot{f} = \dot{z} k_{\text{eff}} \tag{3.6}$$

---

[1]The first condition is of only minor importance because the effect of force loading on the probability of another bond breaking in the range between zero and some low force is rather small, except when using very soft cantilevers. The second condition is much more important, because the effective force-loading rate depends on the molecular compliance and therefore will be altered by the breaking of a bond; we return to this question later.

where k_{eff} is defined as

$$k_{\text{eff}} = \left(\frac{1}{k_c} + \frac{1}{k_{\text{mol}}} \right)^{-1} \tag{3.7}$$

with k_c the spring constant of the lever and k_{mol} that one measured with the slope of the peaks for the group.

To obtain the probability distribution (PD) of the unfolding forces, we have used the Kernel Density Estimation (KDE) method [38] (a discussion to this method can be found in "Appendix"). This technique is equivalent to a histogram but is more precise, giving a smooth representation and converging faster than the histogram towards the real probability distribution function $p(N, F)$.

The different groups of peaks which have to be fitted are labelled by their order N. The values of and uncertainties in the kinetic parameters for each group are deduced from the bootstrap method [39], which requires creating subgroups of the main groups by using the same amount of data with resampling (a representation of the method is shown in Fig. 3.2). For every main group of order N, we have created 500 subgroups, and for each of these we have used the KDE to create a PD which we have fitted with Eq. (3.4) by setting N equal to the main-group index. From the fit of each subgroup, we obtain values for k_{off} and Δx; the final values of these two quantities for each original group is their average over all the subgroups, with their uncertainty taken from the standard deviation.

3.1.2 Experimental results

A total of 190 force curves with at least three clear unfolding events were obtained in the experiment. The distribution of the number of unfolding events on one curve is represented in Fig. 3.3, where curves with six unfolding events are the largest set. To prevent any additional uncertainties, we have thus limited our data analysis to the set of 27 curves with exactly six unfolding peaks and that present a clear and strong detachment peak (clearly stronger than the unfolding peaks).

It is also clear from Fig. 3.3 that our sample is not monodisperse in the number of modules, and that a significant number of molecules have more than 10 modules. This is probably due to the fact that the protein is dimerized. Additionally, the broad distribution of the number of unfolding peaks is a consequence of the circumstance that the molecule is picked up by the tip at random positions along its length.

For a given regime of load, where the conditions of local loading rate are not too different from peak to peak[2], it is possible to linearize the system and consider it like a string made by two springs of different nature. This can be express in theoretical way, by the equation 3.8 with which we interpret k_{mol}. This approach allows us to give an estimation on the total stiffness of the folded

---

[2]This condition has a major importance. In fact, when the load is different from peak to peak it is possible to observe a non-monotonic dependence on the position. This is due mainly to the change of stiffness in the chain rater than the change on the statistics.
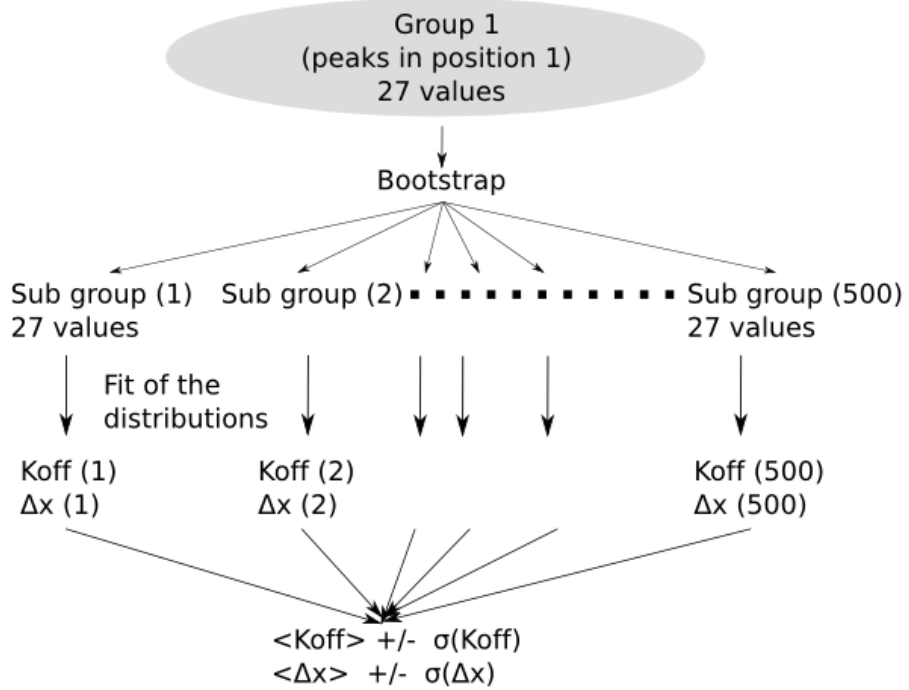
Figure 3.2: Schematic representation of the bootstrap method used to obtain the values of k_{off} and Δx for each set of peaks. This example is for peaks at position 1, meaning that a fit of the distributions requires the use of Eq. (3.5) with $N = 1$.

modules and of the unfolded segments. However this approximation has its limitations: it is known that the stiffness near a peak depends on the kinetic parameters of the proteins itself and on the contour length of the chain. We also know, however, that the stiffness of a long rod can be obtained by the derivative of the WLC model.

$$k_{\mathrm{mol}}(x) = \left(\frac{x}{k_{\mathrm{module}}} + \frac{N - x}{k_{\mathrm{unf}}} \right)^{-1} \qquad (3.8)$$

where "x" represents the peak order, k_{module} represents the spring constant of a single monomer and k_{unf} is the spring constant of a module after the unfolding. If we now plot the average slopes for each group of peaks versus the peak order we obtain the graph 3.4 that gives, when applying the equation 3.8, values of: $k_{\mathrm{module}}(0.121 \pm 0.006)N/m$ and $k_{\mathrm{unf}}(0.0676 \pm 0.003)N/m$.

We have to underline that these values are correct only for the forces near the rupture point because all the components behave like a non linear springs where the "k" is a function of "F".

The average values of the most probable unfolding forces and of their stan-
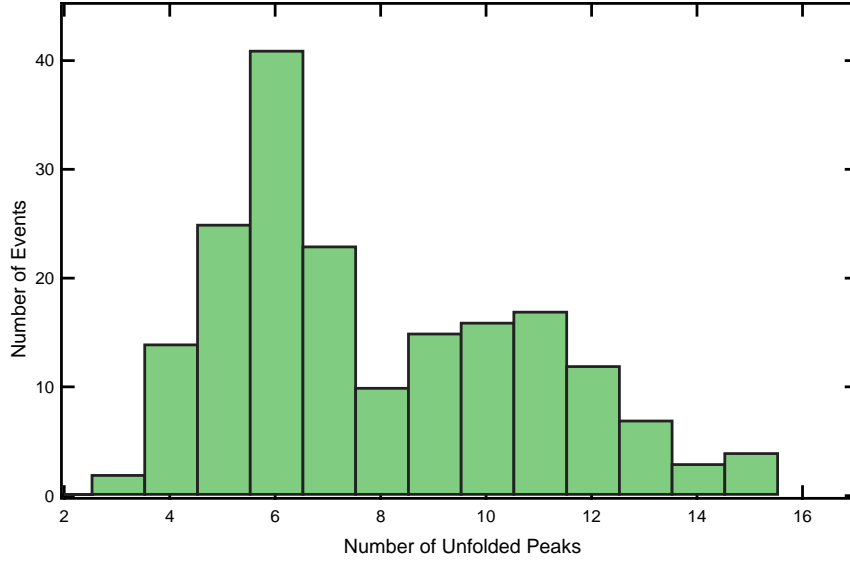
Figure 3.3: Histogram of the number of observed unfolding peaks per curve.

dard deviations, extracted separately for each N, allow a direct test of Eq. (3.5), which predicts a linear dependence of the most probable unfolding force $F(N)$ on $\ln(N)$.

Rewriting eq. 3.5 we obtain eq. 3.9.

$$
\begin{aligned}
F_{\text{rupt}}(N) = \frac{k_B T}{\Delta x} \ln \left(\frac{\dot{f} \Delta x}{k_{\text{off}} k_B T} \right) - \frac{k_B T}{\Delta x} \ln(N) = \\
= F_{\text{rupt}}(1) - \frac{k_B T}{\Delta x} \ln(N)
\end{aligned}
\tag{3.9}
$$

which is the quantity represented in Fig.3.5. Their relationship is indeed very well described by a single straight line, and from the fitted linear dependence we extract the values $k_{\text{off}} = (2.2 \pm 0.3) s^{-1}$ and $\Delta x = 1.15 \pm 0.12$ Å.

We have also studied further characteristics of the multimodular protein unfolding process by applying two different options to process the data. The first option is to find the values of Δx and k_{off} separately for each peak $N = 1, 2, \ldots 6$ by fitting the experimental unfolding force probability distribution to the N-dependent theoretical probability distribution given by Eq. (3.4), to calculate the corresponding weighted averages; by "weighted" is meant that the uncertainties in the values obtained during the fit are accounted for properly during the averaging. This leads to

$$
k_{\text{off}} = (0.37 \pm 0.14) \ s^{-1}, \qquad \Delta x = (1.5 \pm 0.1) \ \text{Å},
$$
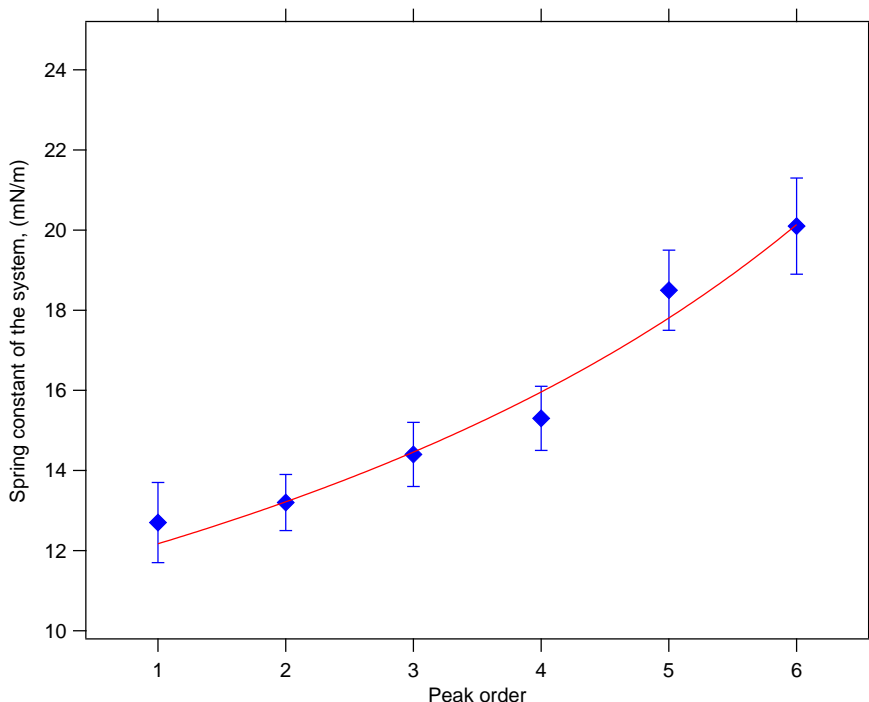
Figure 3.4: Measured average stiffness for different groups of peaks. The elasticity follow the eq.3.8.

and thus to a value of Δx rather close to that extracted from the analysis of the data with Eq. (3.9), as shown in Fig. 3.5.

The second, and conventionally used, option proceeds from an "indiscriminate" fit (one which ignores the order of the peaks) of the experimental unfolding force probability distribution with the form given in Eq. (2.16). This approach, which can be referred to as "single-peak interpretation," results in noticeably different value for k_{off}

$$k_{off} = (4.7 \pm 0.14)\ s^{-1}, \qquad \Delta x = (1.19 \pm 0.08)\,\text{Å}.$$

The values of k_{off} and Δx for each N are plotted respectively in Figs. 3.6 and 3.7. The weighted-average approach is clearly compatible with the values of both quantities for each individual peak, whereas the single-peak interpretation clearly overestimates k_{off}. Figure 3.9 compares the measured dependence of the average unfolding force on N with the theoretical prediction: the values of k_{off} and Δx obtained by the above two data-processing methods were substituted into Eq. (3.5) to calculate the most probable unfolding force as a function of N. Here the error bars for the "weighted-average" and for the "single-peak-interpretation" approaches represent the extremes of the forces obtained using the parameters $(k_{\text{off}} + \delta k_{\text{off}},\ \Delta x + \delta \Delta x)$ and $(k_{\text{off}} - \delta k_{\text{off}},\ \Delta x - \delta \Delta x)$. For the
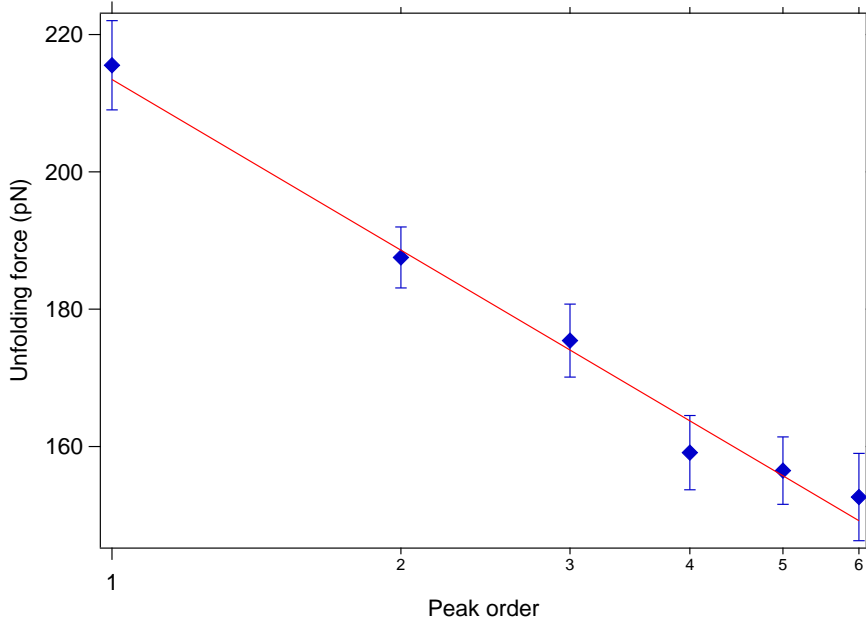
Figure 3.5: Unfolding force versus logarithm of the peak order. The straight line is a fit using Eq. 3.9. The horizontal axis is logarithmic.

experimental data, the error bars represent the standard deviation of the average. From Fig. 3.9, it is clear again that the "weighted-average" interpretation provides a qualitatively better description of the experimental results.

3.1.3   Further comparison with Monte Carlo simulation

Another interesting detail, that arises from this fit, concerns the spring constant of the Monte Carlo simulations that have to be used to reproduce the data. We have compared the experimental data with three different Monte Carlo simulations. The simulations used the same kinetic parameters obtained from the weighted method introduced in the previous section, the only difference was the spring constant of the cantilever. In the first simulation we took the spring constant equal to the average of the spring constants plot in Fig.3.4, in the second we set the spring constant equal to the spring constant of the cantilever, and in the last one we have dynamically change the spring constant of the cantilever to take into account the folded modules according to eq. 3.10, where 0.112 is the spring constant of the cantilever and 0.162 is the estimated stiffness of a folded module.

$$k_c = ((1/0.112) + (N/0.162))^{-1} \qquad (3.10)$$

The results of these three Monte Carlo simulations are show in Fig.3.8.
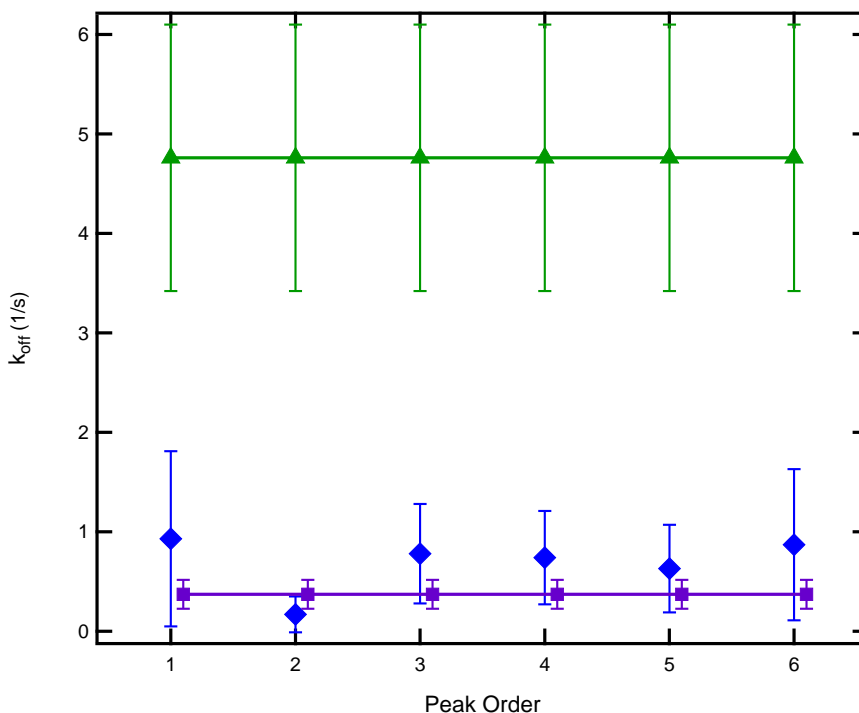
Figure 3.6: Fit of k_{off} for the different peaks. Green triangles are the single peak interpretation. Blue diamonds are the experimental values for each group. Violet squares are the weighted averages.

Here we introduce the following notation that is useful for the following discussion. We define "MC-LL" (Monte Carlo Low Load) the Monte Carlo simulation done with the average loading rate obtained from the slope of the peaks combined with the stiffness of the cantilever. "MC-HL" (Monte Carlo High Load) the simulation done with the stiffness equal to the stiffness of the cantilever. "MC-DY" (Monte Carlo Dynamic) the simulation done with the stiffness that change after the unfolding of each folded module.

In Fig.3.8 we can clearly see that the MC-HL is coherent with the values of the real data for low peak order, however the general behavior is not monotonic. MC-LL is instead coherent with the central and high peak order values and its behavior is monotonic.

This can be justified considering the object that the Monte Carlo simulates. For a certain peak we have the following: the stiffness before the rupture point is given by the WLC function derivative (where the contour length is given by the sum of the contour length of the unfolded modules) composed with the stiffness of the cantilever and, the peak force is given by the statistics of not-yet unfolded modules.

In a real experiment the stiffness before a peak rupture is given by the
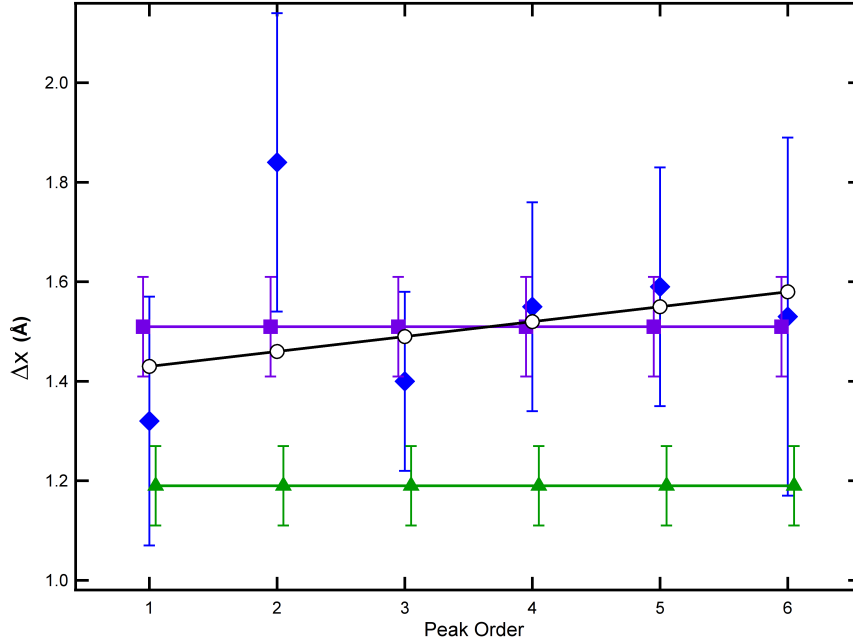
Figure 3.7: Fit of Δx for the different peaks. The black line is obtained from a linear fit, which with $y = bx + a$ yields $a = 1.4$ Å and $b = 0.03$ Å/module with $\chi^2 = 2.11$. Violet squares are the weighted average. Blue diamonds are the experimental value for each group. Green triangles are the single peak interpretation.

stiffness of the polymer unfolded composed with the stiffness of the not-yet unfolded modules composed with the stiffness of the cantilever. Assuming that the WLC derivative mimic the stiffness of the unfolded modules we see that we are still neglecting the contribution of the stiffness of the folded modules.

This justifies the results obtained inf Fig.3.8, in fact in the MC-HL the load at the low value peak order is correctly approximated by the composition of the spring constant of the cantilever and the derivative of the WLC where the contour length is given by the sum of the contour length of the unfolded modules. At high peak order the MC-HL fails because in such region the contribution of the folded modules is neglected.

The MC-LL is not precise for low peak order, in fact for this region the WLC already decrease the loading rate (due to an high contour length) and, substantially, it is like if we are using a much compliant cantilever.

The MC-DY is instead is in good agreement with the data, in fact all the ingredient are present in this kind of simulation. At high peak order the loading rate is moderated by the stiffness of the folded modules, instead at low peak order the compliance is correctly given by the WLC. Differently from the MC-LL, at low peak order, we do not underestimate the loading because the most important contribution is given just by the WLC.
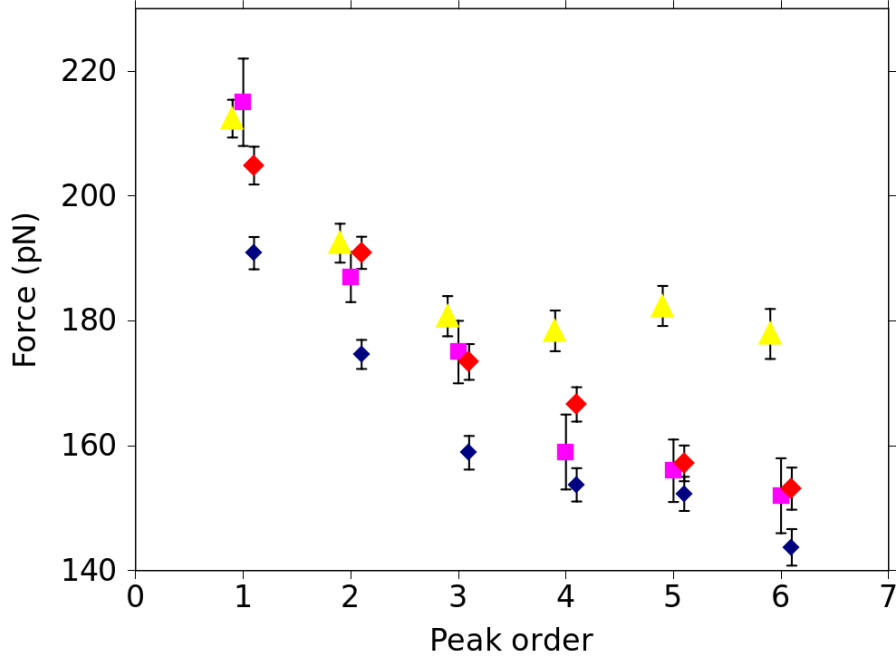
Figure 3.8: A comparison between different Monte Carlo simulation and the real data. Violet square represent the data caming from GB1. Blue diamond came from the simulation with the spring constant equal to the average spring constant of the system. Yellow triangles came from the simulation with the spring constant equal at the spring constant of the cantilever. Red diamond came from the simulation where the stiffness of the cantilever have been dynamically change.

These results make clear that, during a Monte Carlo simulation, it is important to control the current stiffness of the object that is pulled, correcting the spring constant of the cantilever "on-the-run" after each unfolding.

3.2   Conclusion on the kinetic model

The analysis that we have introduced in this section shows how the mechanical unfolding of multimodular homomeric proteins composed by "N" modules can be better studied focusing on the hierarchical behavior in the force curve.

This hierarchical behavior has been explained using the combined statistic of the modules.

We have extended the Bell-Evans theory to interpret the data and consider the effect of the presence of multiple identical modules on the unfolding forces.

The model works under the assumption that, after a module unfolds, the force loading begins again from zero. Such a modification results in a new form for the unfolding force probability distribution [Eq.3.4], which for a multimodu-

lar protein containing N not-yet-unfolded modules predicts most notably a linear dependence of the most probable unfolding force, $F(N)$, on $\ln(N)$ [Eq. (3.5)].

In the previous sections we showed how the model proposed can be applied to the unfolding of GB1 multimodular proteins with $N = 1, 2, \ldots 6$, and can be used to extract the kinetic unfolding parameters.

We demonstrate that results based on the fit of an experimental unfolding force probability distribution with the new theoretical one, and thus on graphs of $F(N)$ as a function of $\ln(N)$, are in good agreement.

Moreover we found that the elastic behavior of the modules and of the linkers near the rupture points follow the behavior of a chain of springs of different nature.

Of course some approximations have been done. In this approach we neglect the combined probability for multiple simultaneous unfolding, as well as a refolding probability which may be important in some cases (especially for experiments with proteins having larger number of modules). However we found no evidence for their relevance in our data set. In a similar vein, the use of Bell-Evans theory may in some instances not be as well justified as it is here.
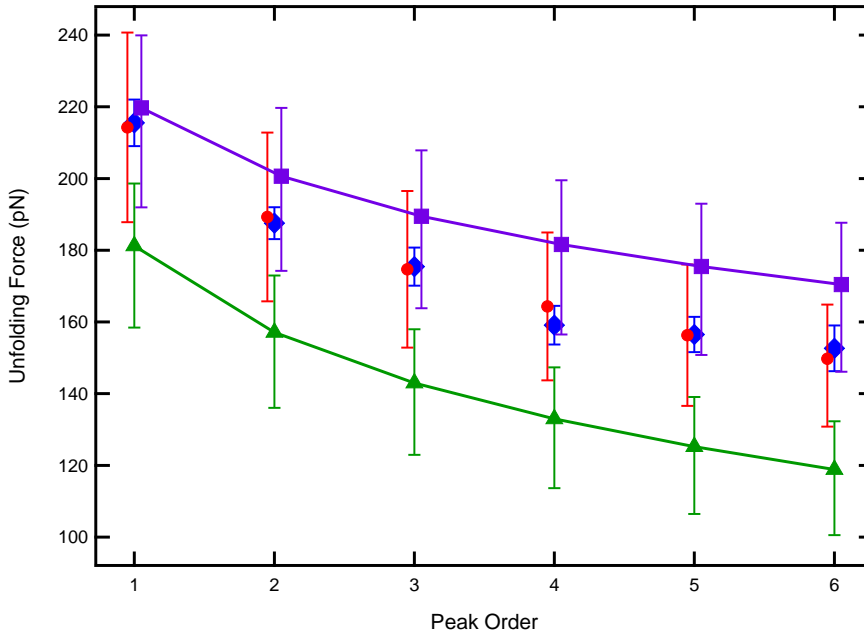
Figure 3.9: Measured unfolding force compared with the force calculated from Eq. (3.5) using values of k_{off} and Δx from Figs. 3.6 and 3.7, shown as a function of the peak order. Blue diamonds are the experimental values (the errors are the σ_{mean}). Violet squares are the distribution method that take in account the "N" dependence. Full red circles are the linear fit with eq. 3.9. Green triangles are the single peak interpretation.

3.3   A phenomenological point of view

As we have seen in the previous section, due to the identity of the modules it is not expected that the forces associated with the unfolding event are drawn from the same statistical distribution (eq. 3.4 clearly define the relations between different distributions), however a relation between the distribution of different events should exist. Furthermore, one may envisage to use only the limited information contained in the experimental distribution of one single group of unfolding forces to predict with high accuracy the average unfolding forces of all other groups. To the best of our knowledge, this last issue has not been addressed before and therefore we sought to investigate it.

The problem is here attacked at two levels. First we adopt a simplified analytical scheme, which implicitly relies on a standard kinetic model for the unfolding of the protein moduli (Evans' theory).

This method, which builds on the results of the previous section, combines a transparent analytical formulation with simplicity of implementation and use. Yet, the simplifying assumptions which allow for the exact analytical treatment of the model come as a disadvantage, since the predicted probability distributions for the unfolding forces of the various peaks can be significantly different from the measured ones.

This limitation can be overcome by using a novel alternative, and more general phenomenological approach which is introduced and discussed later. This scheme, based on the bootstrap statistic and termed backcalculation, is parameter-free and does not rely on any specific kinetic model. The method merely uses the probability distribution of forces associated to one of the unfolding events (the first, or the second etc.) and predicts the distribution of forces of all other events. The method is validated against data obtained from stochastic simulations (both Langevin and Monte Carlo) and from AFM experiments carried out on multimodular $GB1$ constructs. In all cases the average forces associated to any unfolding event are well predicted by the backcalculation and the deviations from the experimental measured values are of only 5pN. This quantity is smaller than the uncertainty typically associated with experimental estimates for protein unfolding forces.

3.3.1   Experiment

The experimental procedure that we have applied is exactly the same as for the kinetic study and can be read in the previous section. The curves have been again selected choosing only those that presented a clear detachment peak. Also in this case the most prominent group was that of the curves with six unfolding peaks. All the curves have been analyzed with "Hooke" and, like in the previous protocol, each peak has been put in a separate group depending on its position in the force curve.

3.3.2 Numerical simulations

Two different computational approaches, namely Monte Carlo and Langevin simulations, were used to study the mechanical unfolding of multimodular protein constructs. In both cases, the pulled construct is assumed to be anchored at one end while the other is pulled at a fixed speed. The end-to-end distance of each protein module projected along the pulling direction, x, is used as an effective order parameter to describe the module state. This corresponds to considering the system as being effectively one-dimensional, as in the sketch of Fig.3.10a. This is a good approximation since, for the typical unfolding forces at play in our experiments, the end to end distance of our construct, as obtained from a worm-like-chain model, is expected to be almost equal to its contour length, so that fluctuations in the y and z directions can be neglected.

Depending on the value of the end-to-end separation, each module is considered as being folded (F) or unfolded (U); these two states are separated by a barrier of potential energy, whose height is modulated by the applied tensile force. The effective potential energy, $U(x)$, is modeled explicitly in the Langevin scheme, where one integrates the stochastic equation of motion for each of the tethered modules in the construct that is being pulled. By contrast, no explicit representation of the construct is considered in the Monte Carlo approach. The latter, in fact is employed to model the succession of discrete unfolding events occurring at force-dependent rates.

The two methods clearly embody rather different strategies for simulating the stretching experiments and, also in view of the different parameters used in the corresponding stochastic simulations, are useful to probe the generality and transferability of the "backcalculation method" (BC) proposed here.

Langevin simulations

With reference to the sketch in the inset of Fig. 3.10a, the anchored end of the construct is located at $x_0 = 0$, while the other end (x_4 in the sketch) is attached to the moving AFM tip. The latter is modeled as a Hookean spring (x_4–x_5 in the sketch) with spring constant k_c=0.01 N/m. Each protein module behaves as an anharmonic spring; the associated free energy profile, $U(x)$, is shown in Fig. 3.10a and described by the following expression

$$U(x) = \frac{A}{L_c - x} + \begin{cases} \frac{1}{2}K_F(x - x_F)^2 \text{ if } x < x_1 \\ U(x) = \Delta G_{FT} - \frac{1}{2}K_T\left(x - x_T\right)^2 \text{ if } x_1 < x < x_2 \\ U(x) = \Delta G_{FU} + \frac{1}{2}K_U(x - x_U)^2 \text{ otherwise} \end{cases} \quad (3.11)$$

The model parameters are chosen consistently with the overall shape of the potential energy landscapes typically found in proteins [40] and are provided in the caption to Fig.3.10. In particular, the contour length of each module is equal to the nominal contour length of $GB1$, L_c =18 nm, the reference end-to-end separation of the folded state is x_F=4 nm and its distance to the transition state is $\Delta x = 0.5$ nm.

In multimodular protein constructs, each protein module is connected to the next via a short peptidic linker of length 1.5nm. To keep at a minimum the number of parameters in the model, we described these linkers, – which clearly do not undergo any transition upon stretching – by unfolded protein moduli. To do so we initially prepare the pristine construct as a succession of folded modules, with initial end-to-end separation equal to x_F, intercalated with unfolded modules, with initial end-to-end separation equal to x_U.

The potential energy barrier separating the F and U states is sufficiently high that an initially unfolded module will not spontaneously refold over the short timespan of the model stretching experiment.

The total potential energy of the homomeric module chain composed by n protein modules, m linkers and the AFM tip is given by

$$H(x_1, x_2, ..x_n, x_{n+m+1}) = \sum_{i=1}^{n+m} U(x_i - x_{i-1}) + \frac{1}{2}k_c(x_{n+m+1} - x_{n+m})^2 \; .$$

(3.12)

The time evolution of the key construct positions, $x_{i=1,...n+m}$ follows the overdamped Langevin dynamics:

$$\gamma \dot{x}_i = -\frac{\partial H}{\partial x_i} + \eta(t)$$

(3.13)

where $\gamma = 4.4 \cdot 10^{-5}$ pN s/nm is the friction coefficient appropriate to yield (according to Kramers' theory) a spontaneous unfolding rate (at zero applied force) equal to $k_{off} = 10^{-2}$ s^{-1}. $\eta(t)$ is a Gaussian white noise term with zero mean and variance equal to $2k_BT\gamma$ (k_B is the Boltzmann constant and $T = 300$ K being the temperature). Notice that the derivative of the potential U (entailed by the derivative of H) is not continuous in x_2.

The stochastic equations of motions were integrated numerically with a time step of 1ns. After an initial equilibration the position of the AFM tip, x_{n+m+1} is moved at constant velocity, $x_{n+m+1}(t) = v\,t$, with $v =$500 nm/s. This velocity value is commonly employed in stretching simulations and falls in the typical range of pulling velocities used in experiments[8]. The typical time-span required to unfold all the $n = 6$ moduli in the constant-velocity simulation was 0.25 s.

The force/extension curve of the system is obtained by recording the restoring force experienced by the AFM tip, $f = k_c(x_{n+m+1} - x_{n+m})$, as a function of the AFM tip position, x_{n+m+1}, as shown in Fig. 3.10b. Several hundred such curves were collected and analysed with "Hooke" after performing a time average over windows of duration 0.15 ms to mimic the finite time-resolution of a typical experiment.

Monte Carlo simulations

As anticipated at the beginning of the section, the Monte Carlo approach (here implemented as in refs.[8, 41] and explained in the chapter 4) provides as phenomenological approach to the kinetics of mechanical unfolding. The advantage

(b)
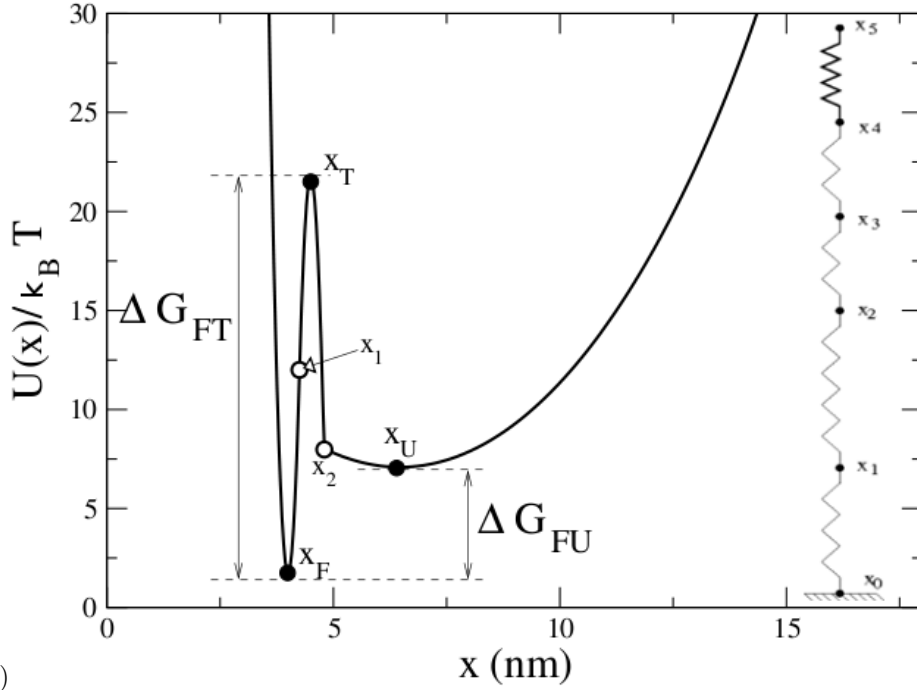
Figure 3.10: (a) Insert: The protein model used in Langevin simulations. (b) Illustration the anharmonic spring potential for one module of the construct in the Langevin simulation, see Eq. 3.11. A large value of A, equal to $100 \, pN \cdot nm^2$, was used in Eq. 3.11 to enforce the constraint that the modules cannot be stretched beyond the nominal contour length, $L_c = 18nm$. The reference end-to-end separation of the folded state, x_F is set equal to 4 nm and the the end-to-end separation between the folded state and the transition (T) state, $\Delta x = x_T - x_F$ is 0.5nm. The reference end-to-end separation of the unfolded state, x_U, is equal to $(L_c - x_T)/2$. Consistently with what established in previous studies, the barrier separating the folded and transition state is set equal to $\Delta G_{FT} = 20 \, k_B T$ while the one between the folded and unfolded state has the value $\Delta G_{FU} = 5 \, k_B T$, with the temperature T being equal to 300K (i.e. $k_B T$ 4.2$pNnm$). For simplicity, the curvature, K_T is set equal to K_F. The value of K_F was, in turn set equal to $4\Delta G_{FT}/(X_T - X_F)^2 = 1344pN/nm$ so to ensure the continuity of the potential and its derivative at the midpoint $x_1 = (X_F + X_T)/2$ where the first two parabolas in Eq. 3.11 meet. The value of k_U was set to be much smaller than K_F, $K_U = K_F/500 = 2.69$ pN/nm. The value of x_2 was finally obtained by the requirement of continuity of the potential. To avoid an excessive parametrization of the model, flexible linkers in the construct are described as unfolded protein modules.
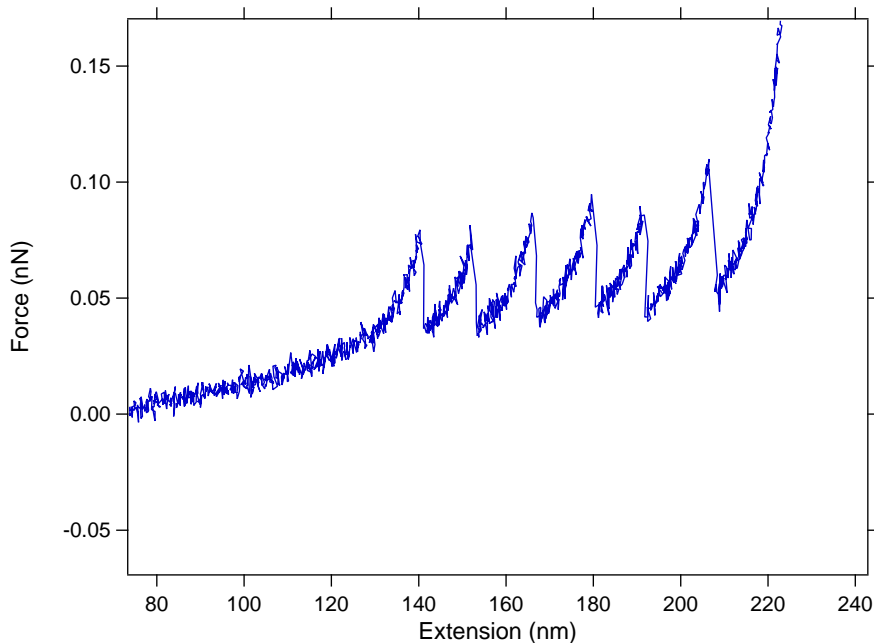
Figure 3.11: A curve obtained with the Langevin simulation with the configuaration: UFUFUFUFUFUFU.

of its transparent formulation is balanced by the highly simplified nature of the model. In particular, by contrast with the Langevin modeling of biopolymers stretching employed here and in other approaches [42], no explicit representation of the module constructs is considered and the linkers are not accounted for. In addition the pulling action is assumed to act equally on all the n modules causing the same steady increase of the end-to-end separation for each of them. Notice that because of the limited sound velocity in the chain this condition is only approximately realized in Langevin schemes and experiments (where other effects such as viscosity can be at play). In any case, the lower the pulling rate the better the approximation is expected to be.

Within the above assumptions, the end-to-end distance (equal to zero at the initial time $t = 0$) of each one of the n modules at time t is equal to $(vt - \frac{F(t)}{k_c})/n$. In this study we considered $n = 6$ and $v = 500$ nm/s and the effective spring constant of the AFM tip is set to k_c=0.01 N/m, as for the Langevin simulation. The instantaneous force experienced by each module is computed from the theoretical force-extension curve, $f_{WLC}(x)$ of an equilibrated WLC with contour length $L_c = 18$ nm (appropriate for $GB1$) and a persistence length l_p=0.4 nm. The progressive loading of the modules is followed at time increments of duration $\Delta t = 1.6*10^{-5}$ s. At the (discrete) time t the probability that one of the modules yields and becomes unfolded is computed within the

Evans's approximation[22]:

$$p(t) = k_{off} \exp(\frac{f_{WLC}(vt) \cdot \Delta x}{k_B T})\Delta t \qquad (3.14)$$

where k_B is the Boltzmann constant, $T = 300K$ is the system temperature. The effective values $k_{off} = 0.11 \, s^{-1}$ and $\Delta x = 1.44$ Å are obtained from a fit of the experimental data using Evans' theory as in ref. [43]. The fitting procedure ensures that the unfolding forces fall in a range similar to the experimental ones, although a precise match is neither expected nor sought. The associated unfolding force is recorded and the calculation is next repeated with the $n-1$ modules. The statistical distribution of the unfolding forces for each value of n was obtained from 1000 repeats of the MC unfolding simulations.
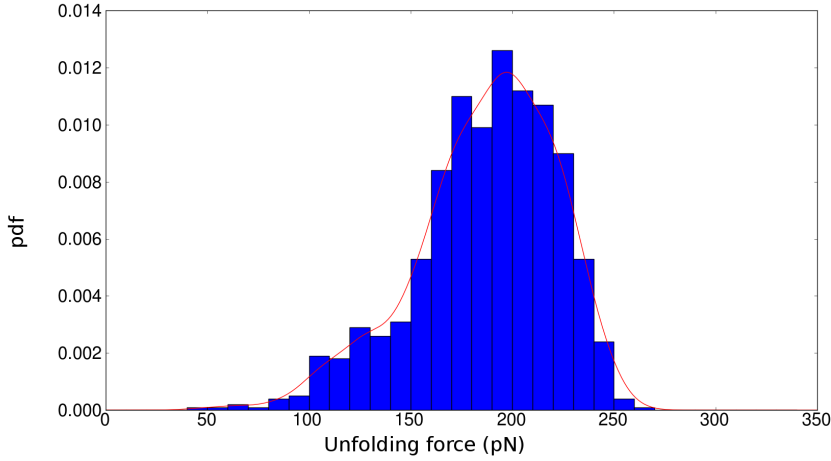
Figure 3.12: Probability distribution of the forces for the peaks near the detachment obtained by Monte Carlo simulation.

3.3.3 Analytically solvable model

Simple analytical expressions for the probability distributions of the unfolding forces, and the associated mean values and variance, as a function of the number of domains, n, can be obtained by introducing a further simplification besides the ones introduced for the Monte Carlo scheme. Specifically, each protein module is treated as an harmonic spring (as in the Langevin approach) rather than a worm-like chain. This solution is equivalent to that shown in Ref. [43], but it is here extended to the analytical calculation of the variance of the unfolding forces.

Let us consider a model construct consisting of N harmonic springs. n of them represent folded moduli and have spring constant equal to K_F while the

remaining $N-n$ have a smaller spring constant, K_U, as appropriate for unfolded moduli. The model construct is subject to the AFM pulling force (the AFM tip is again modeled as a harmonic spring with constant k_c). Because the tip is pulled at constant velocity, v, the tensile force experienced at time t by each construct is equal to:

$$f(t) = \frac{vt}{n/K_F + (N-n)/K_U + 1/k_c} \equiv K_{eff}vt \tag{3.15}$$

where K_{eff} is the effective spring constant of the construct in series with the AFM tip and its inverse decreases with n as $K_{eff}^{-1} = NK_U^{-1} + k_c^{-1} - (K_U^{-1} - K_F^{-1})n \equiv \bar{K}^{-1}(1 - An)$. Here $\bar{K}^{-1}$ is the inverse spring constant of the completely unfolded construct, and A a correction term which describes the dependence of the spring constant on the number of folded modules.

Accordingly, the "survival" probability that any one module has remained folded up to time t is equal to [44]:

$$S_1(t) \equiv \exp[-\int_0^{f(t)} \frac{k_{off}e^{\frac{f\Delta x}{k_B T}}}{vK_{eff}}] df . \tag{3.16}$$

The probability that *all* the n modules have remained folded up to time t, or equivalently up to the loading force $f(t) = vtK_{eff}$, is simply obtained by raising the above expression to the power n,

$$S_n(t) = S_1(t)^n . \tag{3.17}$$

By differentiating S_n with respect to f one obtains the probability distribution, $p(f)$, for the force at which the first unfolding event occurs in a chain of n modules.

The sought expression is

$$p(f) \propto \exp\left(\frac{f\Delta x}{k_B T} - \frac{n(1-An)k_{off}}{\frac{\Delta x\bar{K}v}{k_B T}}e^{\frac{f\Delta x}{k_B T}}\right) . \tag{3.18}$$

where the proportionality factor, containing the normalization of the probability distribution was omitted.

Since the function above is typically non-negligible only for positive f, we can compute its average and variance integrating over $[-\infty, +\infty]$, which leads to the analytical result:

$$\langle f \rangle_n = -\frac{k_B T}{\Delta x}\left[\gamma + \log\frac{k_{off}}{\frac{\Delta x\bar{K}v}{k_B T}} + \log(n - An^2)\right] \tag{3.19}$$

$$\sigma_n^2 = \frac{\pi^2}{6(\frac{\Delta x}{k_B T})^2} \tag{3.20}$$

where $\gamma \sim 0.577$ is the Euler-Mascheroni constant.

We remark here that the variance is independent from the number of folded moduli, n, in the construct. This result is related to the empirical observation that, in typical stretching experiments of a single protein construct, the variance of the unfolding force is largely independent from the loading rate [44].

If the dependence of the spring constant on the number of folded modules can be neglected, the average unfolding force acquires a particularly simple expression:

$$\langle f \rangle_n = -\frac{\gamma}{a} - \frac{\log[b\,n]}{a} \tag{3.21}$$

where the parameters a and b are obtained from the average force and variance for a single, generic $n = n_0$: $a = \pi/\sqrt{6\,\sigma^2}$ and $b = \exp[-\gamma - a\langle f \rangle_1]/n_0$.

Finally, we notice that the expression of eq. 3.18 corresponds to a Gumbel extremal distribution[45] with the "fat" tail extending towards low values of the force, f. Accordingly, the viability of the analytical model to capture the statistical properties of the unfolding forces measured for a given value of n, can be ascertained by checking whether the forces follow the Gumbel distribution. To address this point we employed the Anderson-Darling test and computed the significance level to which one can support the null hypothesis that the data originate from a Gumbel distribution. As customarily, the threshold of 5% statistical significance was used to accept or reject the null hypothesis.

3.3.4 Backcalculation

The previous analytical results rely on a definite kinetic model (Evans' theory) and on the harmonic modeling of the elastic response of the AFM tip and the protein modules. These effects could be included in a more general theoretical framework which, however, would not yield simple analytical calculations.

This difficulty can be circumvented using a simple and physically appealing phenomenological approach, which we term "backcalculation" method and describe hereafter. The method is parameter-free as it relies on the knowledge of the empirical probability distribution of the unfolding forces at one particular value of n_0. This reference distribution can be used straightforwardly to predict the average value of the force and its variance for all other values of n.

The scheme is best illustrated assuming that the reference distribution is the one for $n = 1$, $p_{n=1}(f)$. This distribution is directly obtained from the data gathered in the stretching experiments or from the stochastic simulations, see Fig. 3.12.

In the same spirit of the Monte Carlo simulation and the analytically-soluble model, we assume that the loading rate is sufficiently low so that, at any given time, all modules experience the same instantaneous tensile force applied at their ends, f, and that each of them can unfold independently from the others. We also assume that the stiffness of the construct, defined as the derivative of f with respect to its length, is not dependent on the number of folded constructs n. This is equivalent to considering $A = 0$ in the analytical model, and it's realistic for investigated cases.

Under these assumptions, without resorting to any kinetic model or lengthy stochastic simulations the average unfolding force associated to the nth peak, $\langle f \rangle_n$, is computed by drawing n random numbers distributed according to $p_{n=1}(f)$ and taking the smallest of them as the force at which one of the n modules first unfolds. The average value of the unfolding force $\langle f \rangle_n$ (and its variance) is clearly obtained by repeating the batch force sampling process several times.

At the simplest level, this method is implemented as follows. First one collects N independent experimental (or computational) measurements of the unfolding force at $n = 1$. Then, one draws n elements (with replacement) with equal probability from the sample, retaining the smallest one. This second step is then repeated, and allows to generate a large number of forces which are representative of the n peak distribution. Alternatively, one may use the ordered list of N measurements to construct a cumulative probability distribution interpolated linearly between consecutive measured values. The cumulative distribution is next straightforwardly used (see Chapter 7.3 in [46]) to sample, with the correct weight, the n force values. Describing the process in terms of the cumulative distribution has also the following important advantage. It is possible to exploit the simple relationship of 3.17 (which is based on the assumption of independence and hence valid regardless of the specific underlying kinetic process) to generate data for unfolding forces of the nth peak starting from the data obtained for a peak with a different order, n_0.

In fact, indicating with

$$Q_{n_0}(f) = \int_{-\infty}^{f} df' \, p_{n_0}(f')$$

the cumulative distribution for the unfolding forces of the mth peak, one has that the corresponding cumulative distribution for the nth peak is:

$$Q_n(f) = 1 - (1 - Q_{n_0}(f))^{n/n_0} \ . \tag{3.22}$$

It is important to stress that the above relationships are of high conceptual and practical interest for recovering the distribution of unfolding forces of one peak, say $n = 1$, starting from a peak of higher order, say $n_0 = 2$. A detailed description of this backward extrapolation can be practically implemented in a numerical scheme is provided in the Appendix.

The results discussed hereafter are produced with a more refined method where the probability $p_{n=1}(f)$ is obtained from fitting the histogram of the "raw" force measurements with a convolution of Gaussians using the kernel density estimation (KDE) [38], as shown in Fig. 3.12. Data are sampled according to this distribution using either the cumulative distribution, or the rejection scheme, (see Chapter 7.3 in [46]).

3.3.5   Comparison of the methods

For all the three systems of interest (the GB1 experiment, the Monte Carlo and Langevin simulations) we analysed the data of the force *vs* extension (or

equivalently force vs time) curves. In all three cases the data pertained to the stretching of constructs of $n = 6$ modules and therefore the few curves which did not display 6 force peaks were discarded.

The peak force data for each value of n were considered and used to compute the histograms reflecting the force distribution.

The probability distribution is obtained with a convolution of Gaussians using the KDE method. The resulting normalized distribution of the forces $p_n(f)$ at which a single module unfolds in the Monte Carlo scheme and GB1 experiments are shown in Fig. 3.12. The best-fit Gaussian convolutions were used to obtain a robust estimate of the average unfolding force and its standard deviation at each value of n. The results are provided in Tables and Figs. 3.13, 3.14 and 3.15.

The best-fit distribution for the "last surviving" peak, $n_0 = 1$, was typically used as input for the backtracking and the analytically-soluble methods, so to obtain predictions for the average unfolding forces at all values of n. For the case of highest practical interest, namely the GB1 experiment, the unfolding forces of all other peaks, $n = 2, 3, 4, 5, 6$ were also used (see Appendix).

3.3.6 Monte Carlo simulation data

We start by discussing the application of the method to data generated using the Monte Carlo procedure. Of the three sets of data (experimental, Langevin simulations and Monte Carlo) this set is the one that is expected to be most appropriately captured by the backcalculation. The Monte Carlo scheme indeed builds on the identical kinetic status of all the modules and during this process only the total contour length changes with very mild effect on the loading rate.

By using the $n_0 = 1$ data it is indeed seen in Table 3.1 that the mean values of the predicted and measured unfolding forces are in good agreement for all peaks $n = 2...6$, with differences always below 5 pN. The agreement is readily perceived in Fig. 3.13 where it is seen that the BC data up to $n = 4$ fall within the statistical uncertainty of the MC data and only the forces predicted at $n = 5$ and $n = 6$ present deviations of about 2.5 standard deviations from the MC data.

A more challenging quantity to compare is the second moment of the distribution, that is the variance or, equivalently, the standard deviation. For the latter the agreement is still good. The deviation of the Monte Carlo simulation and backcalculated values, $|\sigma_{BC} - \sigma_{data}|/(\sigma_{BC} + \sigma_{data})$ is typically within 10% and is worst for the last peak, $n = 6$, for which it is 16%.

The results of the analytical model agree with the MC data, that is in turn comparable with the BC one. This is illustrated by the dashed line in Fig. 3.13 which reports the analytical predictions based on the MC data for $n = 1$ (data for this case and other values of n are provided in Appendix). This good agreement is non trivial in view of the fact that the simplified analytical treatment describes the folded protein domains as harmonic springs while the MC data were generated employing a WLC model for each domain. We carried out the Anderson-Darling statistical test for described in the Methods section and es-

tablished that the MC data for $n = 1$ (and higher values too) are compatible with an underlying Gumbel distribution. This reinforces the applicability of the simplified analytical scheme in the model MC context.
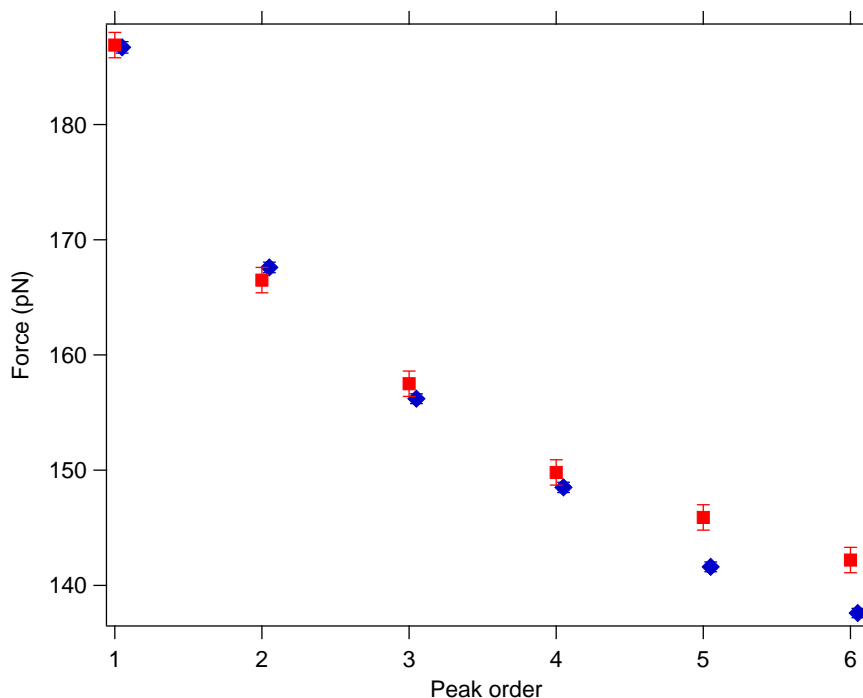
Figure 3.13: Force vs peak order for the Monte Carlo simulation and the BC. In red we can see the Monte Carlo values. In blue the values obtained with the BC.

3.3.7   Langevin data

The same analysis was repeated for the data generated using the Langevin scheme, which contains several differences compared to the Monte Carlo scheme. Specifically, the Langevin scheme does not enforce neither Evans' kinetics, nor the same precise behaviour of all folded moduli in the chain. In addition, it accountes for the presence of model linkers between the folded moduli and finally values of Δx and k_{off} are appreciably different from the Monte Carlo case are used.

As it is visible from Table 3.2 and Fig. 3.13 also for the Langevin context, the performance of the backcalculation method is good and, with the exception of the point for the fourth peak (which compared to the trend of the other data points appears to be an outlier) the average predicted values of unfolding forces are all within about one standard deviation of the Langevin data. As for the MC data, the predicted standard deviations are also consistent with the measured

ones, and the largest relative error, again found for the peak with the largest extrapolation, $n = 6$, is 14%.

As shown in Fig. 3.13 the performance of the analytical model based on the $n_0 = 1$ data is not dissimilar from the one of the backcalculation (the detailed results are again reported in the Appendix). Indeed, also in this context, the Anderson-Darling test indicates that distributions of the unfolding forces are compatible with a Gumbel distribution.
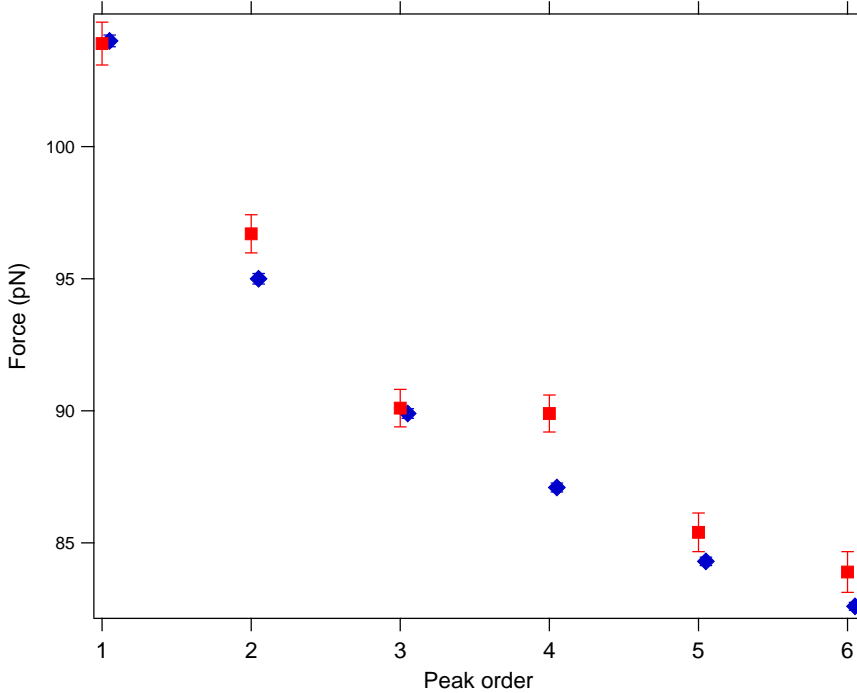
Figure 3.14: Force vs peak order for the Langevin and the BC. In red we can see the Langevin values. In blue the values obtained with the BC.

3.3.8   Experimental data on multimodular GB1

Finally, we turned to the experimental data, which clearly represent the challenge with the highest difficulty. Because of the complex interplay of the several factors that impact on the stretching process and because the pulling rate is not particularly low, it may not be expected a priori that the system unfolding response might be well captured by the backcalculation. Moreover, given the small number of experimental samples, which is equal to 47 per peak, it is not simple to obtain a reference histogram from the experiment or to pin a distribution, even when using the KDE interpolation scheme. Thus, any defect in the starting distribution is consequently amplified by the backcalculation method.

Despite these caveats, the predictive capability of the backcalculation method for the average unfolding forces was found to be very good also in this case. The level of agreement can be appreciated examining Table 3.3 and Fig. 3.15. The increasing underestimation, as a function of n, of the sample standard deviation (predicted from the $n_0 = 1$ peak) can be ascribed probably to the fewer than expected measurements at low forces. This is readily demonstrated by starting the backcalculation from the second peak, $n = 2$, which covering lower values of unfolding forces, can reproduce very well not only the mean unfolding forces at all other values of n, but also the corresponding standard deviations.

In the light of this consideration, the very good consistency of the backcalculation data with the measured distribution is very remarkable, and testifies the robust applicability of the method.

It is particularly instructive to discuss the performance of the analytical method too. Both the average unfolding forces and their standard deviations are not dissimilar from the experimental ones (see Fig. 3.15). However, unlike the cases for the MC and Langevin data, this agreement does not stand a closer statistical scrutiny.

In fact, the Anderson-Darling statistical test indicates that the experimental data do *not* follow the Gumbel extremal statistics entailed by the analytical model. In fact, the null hypothesis for the $n = 1$ peak supported with a confidence level smaller than 1%. The same applies for the $n = 2$ peak as well (in spite of the fact that a more and more pronounced Gumbel-like character is expected as n increases).

The above observation demonstrate the utility of the backcalculation approach in contexts of practical interest. Indeed, the phenomenology of systems such as multimodular constructs of GB1 can be too rich to be well-accounted for by Evans' theory. In such contexts, a good control/prediction of the unfolding forces for varying number of surviving moduli can be made only starting from the phenomenological distribution.

3.4 Conclusions on the phenomenological model

The previous section have presented a systematic investigation of the statistical properties of the forces associated to the first, second etc. unfolding event in a multimodular construct. The phenomenological scheme that we have introduced, named "Backcalculation", use as input the distribution of forces associated to a certain unfolding event to predict the force distribution of all the other events.

As already demonstrated by the kinetic method, it is shown that the standard procedure of analysing experimental protein forced unfolding data by grouping together forces associated to all unfolding events should be replaced by considering separately the events with equal order of appearance.

Secondly, the comparison of the experimental distributions of unfolding forces with the one predicted by standard kinetic models reveals appreciable discrepancies, thus preventing their use as reliable descriptors of the mechanical
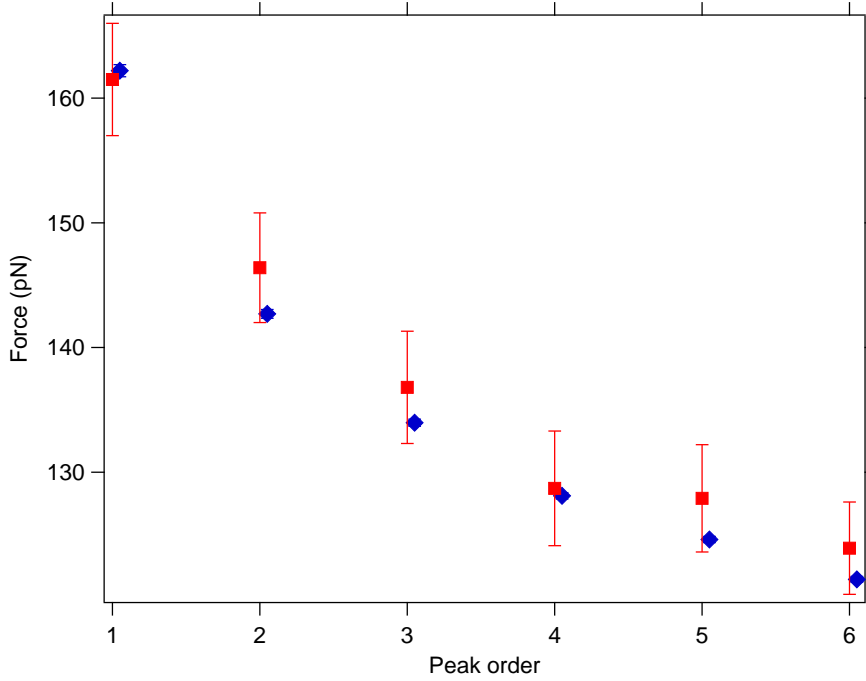
Figure 3.15: Force vs peak order for the experimental data. In red we can see the experimental values. In blue the values obtained with the BC. The BC values have a small horizontal shift to make the graph more readable.

unfolding process. This fact is consistent with previous independent investigations [40].

In addition, the approach has several implications for the design/analysis of stretching experiments of multimodular constructs. First, its simplicity makes the backcalculation particularly appealing as a simple and transparent scheme for the interpretation of experimental data. In this aspect, an interesting application is offered by heterogeneous multimodular constructs, for which the backcalculation can offer a term of reference apt for highlighting composition-dependent modulations of the mechanical response. Second, it offers a simple, parameter-free phenomenological approach to predict the distributions of the various unfolding peaks using a negligible computational effort. In this respect it presents major advantages compared to the more computationally intensive stochastic (Monte Carlo or Langevin) numerical approaches. Finally, it can be applied to design of bio-materials starting from their molecular modular components (e.g. choosing an appropriate number of repeats) with unfolding forces falling in a desired range, or to pre-condition a pulling experiment (choice of pulling speed, stiffness of the AFM tip) so that the mechanical response is profiled with a desired resolution.

Monte Carlo simulation						
n	6	5	4	3	2	1
Average - MC data	142	146	150	157	166	187
Average - $n_0 = 1$ backcalculation	138	142	148	156	168	188
std.dev - MC data	40	35	35	35	37	35
std.dev - $n_0 = 1$ backcalculation	29	31	32	31	33	36

Table 3.1: Average and standard deviations of the unfolding forces as obtained from Monte Carlo simulations and from the backcalculation using data from $n = 1$.

Langevin simulation						
n	6	5	4	3	2	1
Average - Langevin simulation	84	85	87	90	97	104
Average - $n_0 = 1$ backcalculation	83	84	87	90	95	104
std.dev - Langevin simulation	14	14	13	13	14	15
std.dev - $n_0 = 1$ backcalculation	11	12	12	13	14	16

Table 3.2: Average and standard deviations of the unfolding forces as obtained from Langevin simulations and from the backcalculation using data from $n = 1$.

3.5   Further comparisons

An interesting test of our procedure could also be done using the data published in other works where the authors divided the force statistic by peak order of appearance. This have been done to prove that the BC works also for other systems even where the protein studied was not omomeric. Looking at the article of Oberhauser et al. [25] at the Fig. 1 E we can try to apply the BC to the $(I27)_8$ data (open circle). We tried to apply the BC to the data using the peak at highest force (using the Oberhauser notation it is the peak in position "8", for our notation is the peak in position "1"). We have then generated 100 values with Gaussian distribution with $\mu = 227$ pN and $\sigma = 27$ pN (the σ have been taken as large as half of the width of the error bar). The results are reported in Tab. 3.4. As it is possible to see, the difference is at most of 9 pN for the peak near the contact point that is the furthest from the peak that we used for the simulation.

Another interesting article where we can try to apply the BC is Karsai et al. [47], where they analyze the properties of the Cardiacmyosin-bindingprotein-C. This protein is made of eight Ig-like domains and three FnIII-like domains. It is interesting to apply the BC because the protein that Karsai study is not homomeric, still eight of the thirteen modules should show a behavior that could satisfy the BC requirements. Also in this case we obtained the data from the graph in Fig.3 C. As written in the same graph, the data came from four curves. Here it important to see that it is an extremely limited set of data where we are trying to apply our method. Looking at the different sizes of the error bars it is arguable that these bars are the standard deviation of the average.

Unfolding forces for GB1						
n	6	5	4	3	2	1
Average - experiment	124	128	129	137	146	162
Average - $n_0 = 1$ backcalculation	121	125	128	134	143	162
Average - $n_0 = 2$ backcalculation	123	127	131	137	146	160
std.dev - experiment	25	30	31	31	30	31
std.dev - $n_0 = 1$ backcalculation	17	18	19	21	25	35
std.dev - $n_0 = 2$ backcalculation	21	21	22	24	28	36

Table 3.3: Average and standard deviations of the unfolding forces as obtained from experiments on GB1 and from the backcalculation using data from $n = 1$ and $n = 2$. The accuracy in the prediction of the standard deviation is improved when data from $n = 2$ are used.

Comparison							
n	8	7	6	5	4	3	2	1
Average - experiment	186	192	201	197	210	214	218	227
Average - $n_0 = 1$ backcalculation	195	196	198	200	202	206	212	224

Table 3.4: Data obtained from the graph 1 E in [25]. The BC was obtained estimating the average and σ from the peak at highest force.

Looking at value "10" (for Karsai notation) we obtain as average force 129 pN and *sigma* 48 pN. From the theory we also know that the σ is independent from the position and the values of the error bars close to 50 pN are present in the majority of graphs. We can then simulate eight peaks of the eleven peaks present: In fact three peaks belong to the FnIII-like domain and then it is not possible to simulate them with the BC.

Comparison							
n	8	7	6	5	4	3	2	1
Average - experiment	81	78	92	113	132	116	129	152
Average - $n_0 = 1$ backcalculation	78	82	87	94	102	113	130	163

Table 3.5: Data obtained from the graph 1 E in [25]. The BC was obtained estimating the average and σ from the peak at highest force.

In this case we have an error of, at most 30 pN, for the peaks in position 4 (it correspond to the peaks in position 8 for Karsai) but, also looking at the graph, it is possible to say that it could be an outlier. The prediction for the other peaks is quite good and it does not exceed 9 pN. So even in this latter case, where we have a multimeric construct, we have a good prediction of the force peaks. This is a strong evidence that the BC method could be applied to many (even if not all) multimodular construct.

Chapter 4

Monte Carlo for the AFM

4.1   Monte Carlo for multimodular proteins

In a typical AFM pulling experiment a single molecule, fixed to a surface at one end, is pulled by a cantilever tip at other end. What we usually measure is the force as a function of the distance of one end of the molecule from the surface. In the case of proteins (simple or multimodular) we can observe peaks on the force-distance graph, where the force follows closely the well known Worm Like Chain (WLC) model eq.2.6 [48] where the l_c is the characteristic contour length of the molecule and l_p is the persistence length.

Also several kind of polysaccharides molecules (e.g. dextran, cellulose, amylose and others [49, 50, 51]) have been pulled, in these cases we can use the WLC or the FJC model to follow the force extension curve (excluding the non-entropic transition regions).

The theory of Evans [22] has been used to interpret force spectroscopy experiments of protein-protein interaction or protein unfolding to obtain the kinetic parameters that characterize the energy landscape. However we have to remember that, due to the geometry of pulling, the kinetic parameters represent the energy landscape only along the coordinate given by the direction of pulling, so they can take into account only a single, peculiar unfolding pathaway.

This theory can be used for the inverse problem, to generate a force spectroscopy pattern based on the properties of a polymer[41, 52, 53] and its kinetic parameters. This kind of simulations can provide useful informations to the experimentalists and it may turn useful as a test for hypothesis. Moreover the Monte Carlo can be used to fit the force extension curves to find the kinetic parameters, this has been done before with different methods [35].

The main idea of the Monte Carlo simulation for protein pulling is to apply

a force "F" to a chain with a determined protocol. For instance, we will assume that the force follows the WLC function in the case of velocity clamp. But we can also keep the force constant and, in such case, we will have a force clamp. Using the kinetic parameters we can calculate the transition rate for a force "F" as following:

$$\lambda(F) = k \ exp\left(-\frac{\Delta G - F\Delta x}{k_B T}\right) = k_{\text{off}}exp(F\Delta x) \tag{4.1}$$

For the simulations that we are presenting and for the fact that, in our lab we use the velocity clamp techinique, we will present the case for which the force at time "t" follows the WLC.

When we simulate such process we have a slight complication. The WLC formula is a function of "x" and this latter is a function of time and of the bending of the cantilever.

Looking at the Fig.4.1 we can see that we can write "x(t)" as eq.4.2.

$$x(t) = vt - c_b(t) \tag{4.2}$$

The bending of the cantilever depends also on the total force that is felt by the system at the time "t" and can be written with the formula 4.3 where k_c is the spring constant of the cantilever.

$$c_b(t) = F(t)/k_c \tag{4.3}$$
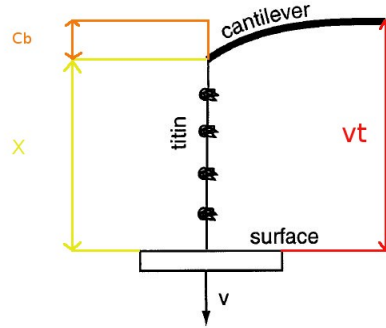

Figure 4.1: Schematic representation of the distances during an AFM experiment.

Inserting equations 4.2 and 4.3 in equation 2.6 we obtain an equation for the force that explicitly depends on time.

$$F(t) = \frac{k_B T}{l_p}\left[\frac{1}{4(1 - \frac{vt - F(t)/k_c}{l_c})^2} + \frac{vt - F(t)/k_c}{l_c} - \frac{1}{4}\right] \tag{4.4}$$

The Eq. 4.4 is a third grade equation in F(t).

In our simulations we did the following substitution: $\frac{vt - F(t)/k_c}{l_c} = a$ (see also the code in Appendix). Then writing $F(t) = (vt - al_c)k_c$ we got (after some rearrangements) the following equation:

$$(4vt - al_c)\frac{k_c l_p}{k_B T} = \frac{1}{(1-a)^2} + 4a - 1 \tag{4.5}$$

Writing $\frac{4k_c l_p}{k_B T} = B$ and recasting we have:

$$a^3(-4 - Bl_c) + a^2(Bvt - 2Bl_c + 9) + a(-2Bvt - Bl_c - 6) + Bvt = 0 \tag{4.6}$$

This equation, using the experimental parameters, can be solved for "a" with numerical methods (the Newton solver was applied in our work) and it gives always a real result for "a". Knowing "a" we can easily calculate the "F(t)".

The equation 4.4 is correct for most of the applications but we need to notice some important details.

As we said in chapter 3, the stiffness of a protein composed by modules under tension in an AFM setup results from the following components: the spring constant of the linker, the spring constant of the folded modules and the spring constant of the cantilever. All these components are connected in series and the final value of the stiffness (near an unfolding point) is given by Eq. 4.7.

$$s_s = (1/k_c + 1/k_l + m/k_f)^{-1} \tag{4.7}$$

The WLC formula takes in account the change of contour length (that means not only an increase of length but also a change in the stiffness of the polymer), but it neglects the contribution of the folded modules to the total stiffness (because they "do not exist" until one of them unfolds and increases the total contour length). This problem can be circumvented by decreasing the spring constant of the cantilever according to the Eq. 4.7 such that $k_s = k_c$.

Once we know the force we can calculate the unfolding probability, but we need to pay attention also to the temporal step of the simulation. At every discrete increase of time, the molecule sustains a force F+dF and an unfolding probability p(F) (eq. 4.8); if we choose a coarse timestep, the MC simulation will not have the sensitivity to explore the "detail" of the energy barrier and for a simple barrier, we will have not enough "attempts" of crossing (or in other words we will not have a good statistic and a good sampling of the WLC function). As reported in [52] we should choose a timestep small enough to have a good statistic - also taking in consideration, of course, that if we decrease the timestep we will increase the duration of the simulation.

If we are pulling a multimodular protein, to simulate it correctly, we have to take in consideration that in the eq.4.4, after one unfolding, the l_c of the value of the unfolded module and the number of the folded module "N" decrease by one. Moreover, at the same time "t", many modules sustain the force and the unfolding probability. This involves a parallel process that has to be taken into consideration. We can still run a sequential experiment, but to correctly express the unfolding probability of the modules in a chain we have to use the eq. 4.8

where "N" represent the number of folded modules. During the simulation, to know if a module unfolds, we have to extract a random number between zero and one at time "t" and we have to compare it with the value obtain by eq. 4.8 (where Δt is the timestep) with the force at time "t".

When the random number extracted is lower than the probability we will have a transition (it can be an unfolding or a shape change), if not the process start again but now with a time "t+dt" ,and correspondingly a higher force.

After an unfolding event the parameters will change as follows:
the contour length become $((l_c)_{new} = (l_c)_{old} + (module)l_c)$ and the number of folded module $(N_{new} = N_{old} - 1$).

$$p(F, N) = N * \Delta t * \lambda(F) \qquad (4.8)$$

With the correct parameters the aforementioned scheme can be easily applied to generate force extension curves or to collect a statistics of unfolding forces for multimodular protein.

4.2   Polysaccharides simulation

The scheme that we have presented can be applied to the polysaccharides with some changes. It is known that, under the condition of load of a typical AFM pulling experiment, the stretching process for polysaccharides is at equilibrium. This fact has a series of implications: the stretching curve exactly corresponds to the relaxation curve and the process is speed independent[54] (see also Rief et Al.[52] where they simulate the extension of a polysaccharide chain). The scheme can be used to understand the similarities between the this kind of chain and a multimodular protein.

One of the main differences between the unfolding of multimodular protein and polysaccharides is that, to simulate an equilibrium process, we have to use a really large k_{on} rate (in Rief[52] a value of $k_{on}/k_{off} \sim 10^6$ was used).

This rate represents the probability to have a reversed transition.

This can be done by introducing in the simulation (see the details in appendix 10) a probability similar to eq. 4.8 but that depends on the number of transitioned monomers and on different kinetic parameters. In particular, to have some effect on the force curve, the k_{on} should be several order of magnitude bigger that the k_{off}. The process of refolding/reverse transition is similar to unfolding. Also in this case a random number is extracted, if this value is lower that the reverse probability we will have a transition. In this case the number of folded modules increases and the contour length is decreased (so this time we will have: $(l_c)_{new} = (l_c)_{old} - (module)l_c$).

4.3   Parallel bond breaking

Another interesting system that can be modelized with the Monte Carlo is the protein-protein interaction. Several studies have been done on this topic [55][56]
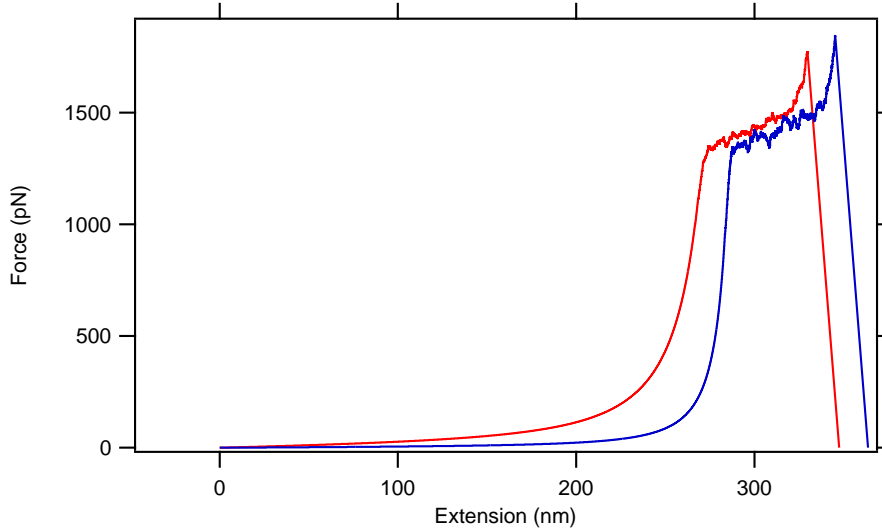
Figure 4.2: A simulation of a force distance curves. The parameters for the two curves were similar, a number of modules of 600, speed 1.33 $\mu m/s$, k_c 0.1 N/m, initial module length 0.44 nm, module l_c 0.1 nm, temperature 300 K, $\Delta x = 0.042$ nm, $k_off = 2.5 * 10^{-5}$ s^{-1}. The difference in the two is only in the value of the l_p, the red curve has a l_p of 0.1 nm the blue of 0.5 nm.

[57][58]. In a typical ligand-receptor experiment the experimentalist functionalizes the AFM tip with the ligand and the surface with the corresponding receptor (or viceversa). This functionalization is performed with various chemical techniques (see also the capther 2) . These procedures allow to study the ligand-receptor interaction using the AFM. However the number of bonds that are formed during the interaction is not controlled.

We can easily affirm that the ligand-receptor interaction is a more challenging subject than the study of multimodular protein. The main reason is the lack of a reference pattern that makes it difficult to discriminate the various event types: one, two or multiple ruptures.

One of the main consequent problem is still how to identify a genuine ligand-receptor interaction in a force spectroscopy experiment. In fact, also in well controlled conditions, the force peaks may be different in term of force/length-/slope. This may be due to the fact that, during the experiment, the probe, functionalized with a ligand/receptor, interacts with more than one receptor/ligand on the surface, resulting in many different types of force peaks.

In some studies it has been tried to classify the force peaks observed with different number of interactions [55], in order to show, with a probability distribution representation, a quantization in the maximum rupture force.

Following a theoretical approach, we found that it is possible to have a quantization, but such effect can be only interpreted with a generalization of the Bell-Evans theory for multiple bonds. Such quantization does not follow a

simple rule where two bonds have just twice the force of a single bond but it is more complicated and can be explained with the model that we are going to expose.

For this reason we are going to introduce two models, simulate several points with a Monte Carlo scheme and then fit the results to see if the theoretical model can recover the values used in the simulation.

4.3.1   Theory for parallel bonds

In a common Monte Carlo simulation for protein pulling we consider a system that is similar to a series of springs that can undergo a short to long transition. However we can think of different models that could be useful to interpret force spectroscopy experiments.

During the studies of the avidin-biotin system we had the necessity to develop an interpretation method that could deal with a large spread in the force *vs* loading rate graph. We have imagined that, during an experiment, when a ligand (or receptor) is on the AFM tip and a receptor (or ligand) on the substrate, we could have multiple parallel interactions (as illustrated in Fig. 4.3).
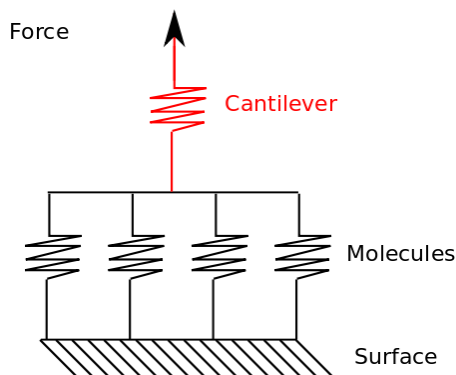
Figure 4.3: In the figure we can see a schematic representation of a system with parallel bonds. The bonds are represented as black springs and the cantilever in red. In such configuration, the bonds divide equally between themselves the force applied through the AFM cantilever.

We have then modeled a system that is composed by a series of parallel springs that can undergo a transition depending on the loading rate. In the simulation, the force is distributed equally among these springs.

Compared to the previous Monte Carlo methods illustrated before, here we do not have multiple peaks, and at the breaking of a bond we do not have any increase in the contour length, however the breaking of a bond influences the process, by decreasing by "1" the number "N" of not yet broken bonds.

When a system of parallel bonds is pulled, the transition rate λ is given by eq.4.9. The difference between eq.4.9 and eq.2.10 is due to the fact that the force is now equally distributed between all intact bonds.

$$\lambda(f) = k_{\text{off}} \ \exp\left(\frac{f\Delta x}{Nk_B T}\right), \tag{4.9}$$

Exactly as for a series of bonds in chapter 3, also in this case we want that during the force load from $f = 0$ to $f = F$ "N" bonds survive. Thus lead to the equation 3.4 saw in chapter 3. Using the same procedure we can obtain the probability distribution and the maximum of the force.

$$p(f) = Nk_{\text{off}} exp\left(\frac{f\Delta X}{k_B T N}\right) exp\left(\frac{k_B T k_{\text{off}} N^2}{\dot{f}\Delta X}\left(1 - exp\left(\frac{f\Delta X}{k_B T N}\right)\right)\right) \tag{4.10}$$

Deriving this equation to find the maximum force with $dp(F)/dF = 0$ easily leads to eq.4.11 that express the most probable force (F_{mp}).

$$F_{mp}(N, \dot{f}) = \frac{k_B T N}{\Delta X} ln\left(\frac{\dot{f}\Delta x}{k_B T N^2 k_{\text{off}}}\right) \tag{4.11}$$

Now if we take in consideration the loading rate, and if we are using a stiff cantilever, we have eq.4.12.

$$k_s = \left(\frac{1}{Nk_m} + \frac{1}{k_c}\right)^{-1} = \frac{Nk_m k_c}{Nk_m + k_c} \tag{4.12}$$

Where k_c is the spring constant of the cantilever and k_m is the spring constant of a molecule. Now we can say that $k_c >> Nk_m$ and thus we can simplify eq. 4.12 and write $\dot{f} = Nk_m v$. If we put this latter equation in 4.11 we obtain:

$$F_{mp} = \frac{k_B T N}{\Delta x} ln\left(\frac{k_m v \Delta x}{k_B T N k_{off}}\right) \tag{4.13}$$

When instead $k_m > k_c$ we have that $\dot{f} = k_c v$ that lead to eq.4.14.

$$F = \frac{K_B T N}{\Delta x} ln\left(\frac{k_c v \Delta x}{K_B T N^2 k_{off}}\right) \tag{4.14}$$

4.3.2 Implementation

The transition rate is given by 4.9, taking into account the time step. The procedure to increase the force is similar to the multimodular protein Monte Carlo simulation, however here we can think to two different schemes to increase the force:

- In the first scheme we can directly use the WLC equation that is independent on the number of ligands receptors. At each time step we can check how many bonds will break. This mean that, during the pulling, if one bond breaks, the function to increase the force (the WLC) will not change

and the cantilever will not feel any relaxation. On the other way around
we can also think that the elongation of the cantilever is much bigger than
the elongation of the molecules. We call this scheme "stiff bonds".

- In the second scheme a prefactor "N" appear at the beginning of the
  WLC equation so we have $F_N(t) = NF(t)$. This latter has been obtained
  from the consideration of the fact that for a system of parallel bonds the
  derivative of $F_N(t)$ should give the stiffness of the bonds like Nk_m. So
  with this last scheme, after the breaking of a bond we should observe
  a relaxation on the cantilever. This scheme was called "soft bonds" or
  "springlike".

4.3.3   Study of a set of systems with different number of bonds

Here we are going to show the behavior of different systems composed by a
different number of bonds. The bond parameters of the various systems are
the same in terms of Δx and k_{off}, the difference between the systems relies on
the number of parallel bonds. For each system we have simulated 10 thousands
points using the "hard bonds" scheme. Each point represent the rupture force
of the last surviving bond. The parameters for the simulations were:

- speed 100 nm/s,

- $k_c =$0.05 N/m,

- l_c 20 nm,

- l_p 0.3 nm,

- temperature 300 K,

- Δx 0.15 nm,

- k_{off} 0.01 s^{-1}.

The first result that we want to show is the Fig.4.4. The results show that
all the systems stay on the same curve and, as expected, increasing the number
of parallel bonds the rupture force increases.

However, when we deal with a real system we observe a large spread of data
that do not stay on a line. Can we take a section of points on a range of loading
rate and see if there is any quantization? To answer to the question we took all
the rupture forces of our simulation from 3.5 nN/s to 20 nN/s and we create
the correspondent KDE.

In Fig.4.5 it is possible to see that in the KDE of all the events there is
a quantization -however, when we approach the higher forces, we start to lose
sensitivity. This latter effect is probably due to the fact that the width of the
distributions increase with the number of parallel bonds, making the process
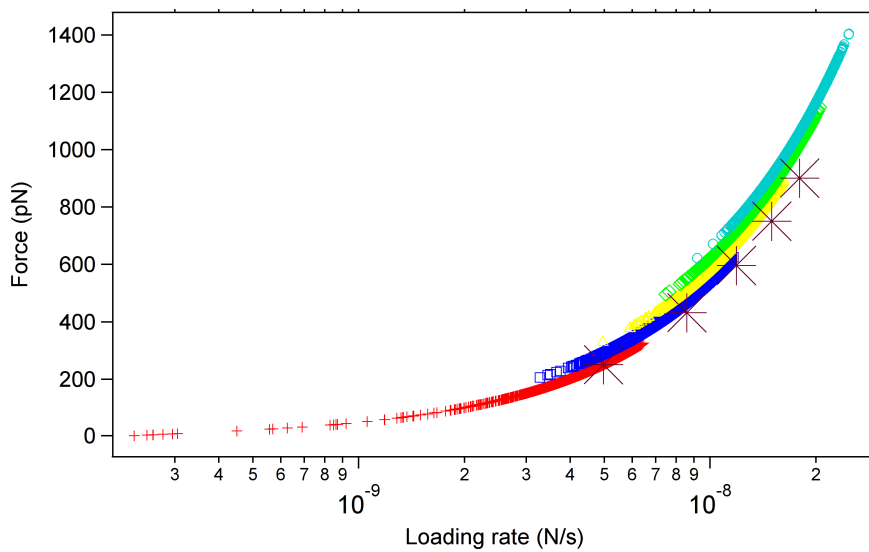
Figure 4.4: Here we can see the results of the Monte Carlo simulations using the hard bond scheme for systems with a different number of parallel bonds. All the points stay on the same curve, an offset was put to better distinguish the various systems. Red cross represent the system with one bond, blue square 2 bonds, yellow triangles 3 bonds, green diamonds 4 bonds and azure circles 5 bonds. The violet stars represent the average values (for rupture force and loading rate) for the various systems. They have been plotted to show the centrality of the distributions of points.
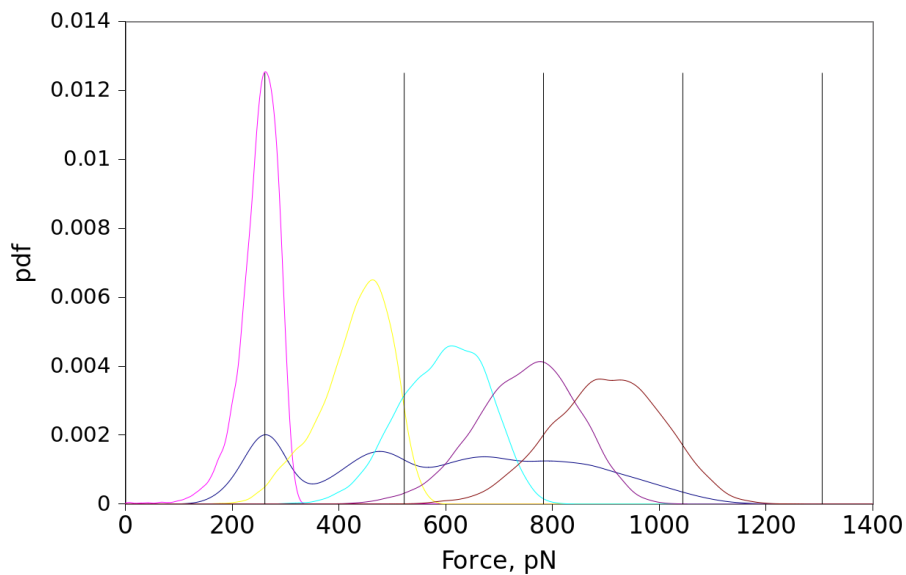
Figure 4.5: In the figure we can see the KDE for various systems. Violet for the system with one bond, yellow 2 bonds, azure 3 bonds, violet 4 bonds, brown 5 bonds. The blue is the KDE of all the points of the aforementioned systems. All the KDE have an area of one, for this reason the KDE of all points has a lower height, compared to the other KDEs. The black vertical lines represent the integer multiples of the maximum force of the system with one bonds. These lines were made to show that the maximum force of the distributions does not have just "n" times the force of the system with one bond. An interesting feature that we can see is the increase of the width of the distributions.

of restoration of the central value of the distribution for highly parallel bonds difficult. It is the opinion of the author that it could be possible to deconvolve such distribution of forces with a numerical/analytical method, however such a method would require much more effort and it has not been developed yet.

4.3.4 Application to ideal MC system

In this section we will try to apply the theory that we have developed to fit simulated data, obtained from a Monte Carlo simulation of a parallel bonds system. The general idea of the procedure is that a set of data follows the equation 4.13, however in this equation we have to stress that k_m is a function of "F". In fact in the simulation we are using the WLC function to increase the force and, using this function, higher forces mean a different slope. A different slope means also a different k_m. The assumption that we use in this fitting is that $k_m(F(N)) \approx k_m(F(1))$, that is that the slope of a force peak, for this kind of interaction, does not strongly depend on the number of bonds. Using this approximation we should see a quantization of the forces in a small region of the force *vs* loading rate graph.

The data were generated using the "stiff bonds" scheme.

In a real experiment, also using the same pulling speed, we get a large spread in the loading rate *vs* force scatter plot. This may be due to different interaction geometries of the ligand/receptor complex, in fact we cannot expect that the interaction between the ligand/receptor on the tip and on the surface to be perfectly vertical.

To model these different geometries, that give rise to different loads on the complex, we have used a random velocity in a range between 5 and 105 nm/s (with a uniform distribution). We have then run 10 thousands simulations where the number of bonds for each one was randomly chosen between 1 and 10 (with a uniform distribution). Each point in the graph correspond to the maximum value for the system generated. The values generated can bee seen in Fig. 4.6. The parameters used were:

- spring constant cantilever 0.050 N/m,

- contour length 20 nm

- persistence length 0.3 nm

- $\Delta x = 1.5$ $\mathring{A}$

- $k_{\text{off}} = 0.01$ s^{-1}

To analyze the results we need to obtain a distribution of probability of rupture forces for 1 to N bonds.

We have assumed that the loading rate at the rupture point of "N" bonds is not "N" times bigger than a single one and we also know that this dependence is logarithmic; as an approximation we can select a relatively small slice in the force *vs* loading rate graph to capture the different behavior of the systems with
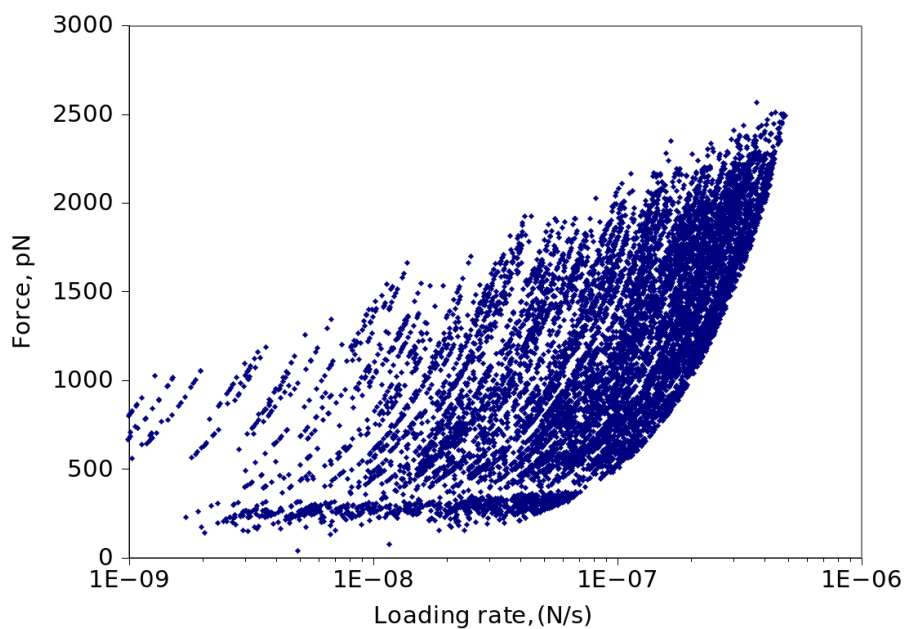
Figure 4.6: Scatter plot for the Monte Carlo simulation of "stiff bonds" with different loading rate and different number of bonds. The parameters of the simulation were: k_c 0.05 N/m, l_c 20 nm, Δx 0.15 nm, k_{off} 0.01 s^{-1}. The number of bonds were chosen, randomly, from one to ten and the velocity from 5 to 105 nm/s.

1 to "N" bonds (however this point needs more investigation). Looking at the scatter plot of the data obtained from the MC we have selected the points in the range of loading rate between 16.5 to 17.5. With these points we have created a KDE and to find the maximum we have done a double differential, the KDE and the differential can bee seen in Fig. 4.7.
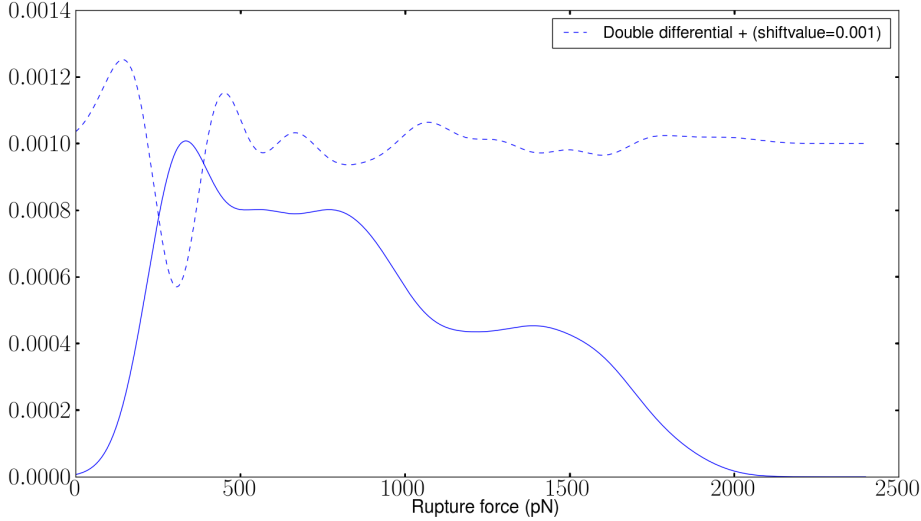
Figure 4.7: KDE of the points in range of loading rate between -16.5 to -17.5 (of the graph 4.6). The maximum of the sub-populations has been identified using the double differential.

Using the graph 4.7 we can now plot the maximums of the various distributions versus the number of bonds that we suppose for each distributions. The correlation between the maximum of the distributions and the bonds has been studied assuming that the maximums depend linearly from the bonds number.

The fit in graph 4.8 give $\Delta x = 1.8$ $\mathring{A}$ and $k_{off} = 0.0015$ s^{-1}. The Δx value obtained is quite close to the value used in the simulation (1.5 $\mathring{A}$, the k_{off} may seem not so good instead (in the simulation $k_{off} = 0.01$ s^{-1}), but we have to remember that this value is always strongly influenced by the error. In fact if we look at the formula for most probable force (eq. 2.17) we can see that the k_{off} appear inside the logarithm when Δx also as multiplicative factor. This mean that even a large change in the k_{off} is diminished by the logarithm. So, obtain a value that is just one order of magnitude different from the k_{off} used is still reasonable.

4.3.5 Application to the Hemmagglutinin New Caledonia *vs* 6'SLN or 3'SLN

The model that we have introduced was then applied to a real system to extract the kinetic parameters. We have analyzed the data coming from the interaction
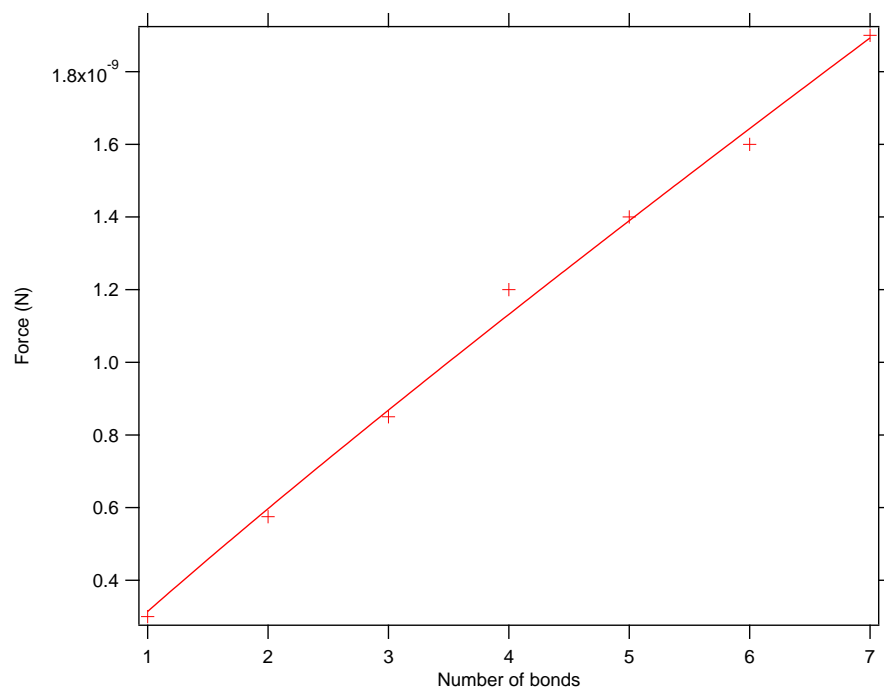
Figure 4.8: Most probable value of the distribution *vs* number of bonds. The points have been obtained from the KDE inf Fig.4.7 and then fitted with eq. 4.13.

of Hemmaglutinin (HA) New Caledonia *vs* 6'SLN (6'-Sialyl-N-acetyllactosamine) and HA *vs* 3'SLN (3-sialyl-N-acetyllactosamine). The first step of the analysis is to create a scatter plot of for the peaks in the logarithm of loading rate *vs* force graph. Using this scatter plot we can then choose a range of loading rates to analyze (depending on the density of data).
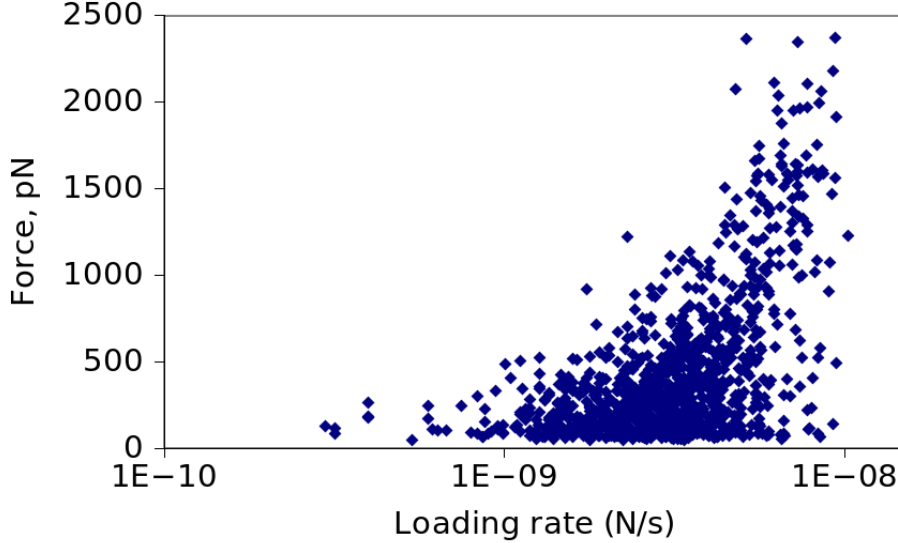
Figure 4.9: Scatter plot for the peaks coming from the experiment with HA New Caledonia *vs* 6'SLN.

In Fig. 4.9 we have chosen a range of loading rate from -20 to -19, the points in this range have been plotted in a KDE reported in Fig. 4.10.

If we try now to plot the maximum that we can obtain from the second derivative of the KDE versus the hypothetical number of parallel bond we get the Fig. 4.11.

The fit in graph 4.11 gives: $\Delta x = 3.3$ $\mathring{A}$ and $k_{off} = 2.7 * 10^{-5}$ s^{-1}, in the fit we have used an average value for the loading rate of -19.25.

We have tried the same approach for the system HA New Caledonia *vs* 3'SLN, the distribution of the points are reported in graph 4.12, the KDE in the range of loading rate -19 to -18 ln(N/s) can be seen in Fig. 4.13 and the graph for the maximum forces *vs* number of bonds is in Fig. 4.14. This latter fit yields the following kinetic parameters: $\Delta x = 6.9$ $\mathring{A}$ and $k_{off} = 1 * 10^{-5}$ s^{-1}

4.3.6   Conclusion

The derivation of the maximum force for a system with parallel bonds was obtained using the Bell-Evans theory. This derivation uses a minimal number of hypothesis and give predictions that are well validated by Monte Carlo simulations. The most probable force for system with 1 to "N" bonds was calculated
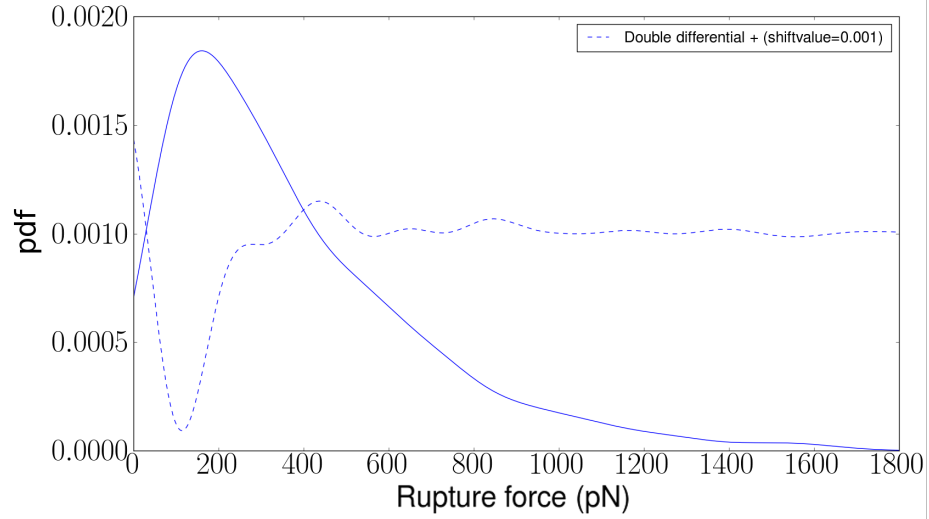
Figure 4.10: Probability distribution of the forces in the range of loading rate between -20 to -19.0 for the peaks coming from the experiment with the HA New Caledonia *vs* 6'SLN. The continuous curve is the KDE for the ruptures forces. The dashed line is the double differential of the KDE. The total number of events used for the KDE were 780.

from Monte Carlo data. These values were then fitted with the model, and we were able to get back the kinetic parameters used in the simulation. However a difference of an order of magnitude is observed in the k_{off} - this is still the most challenging parameter to estimate. Following the multiple parallel bonds model we have supposed that a particular distribution in the force *vs* loading rate graph should exist. Such distribution should reflect the eq. 4.13 (the equation gives the maximum of a distribution for "N" parallel bonds) and so, for a determined range of loading rates we should be able to distinguish different maximums. We have then tried to apply this idea to simulated systems and real systems. For the distribution of maximum values obtained from the simulation we can say that the results are in good agreement with the parameters used to generate the points. For the real data we get reasonable results but there are no others estimations with which we can compare. Moreover this interpretation gives an answer to the distribution of data and to the apparently different force peaks that an experimentalist might see during the experiments.
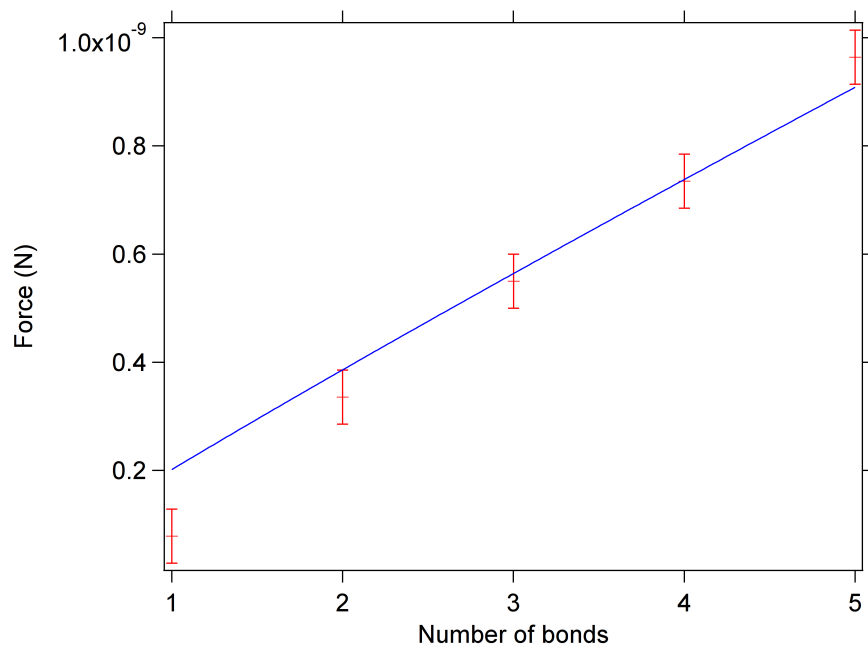
Figure 4.11: Maximum of the distributions *vs* number of bonds for the system HA New Caledonia *vs* 6'SLN obtained from the KDE in Fig. 4.10. The values have been fitted with eq. 4.13.
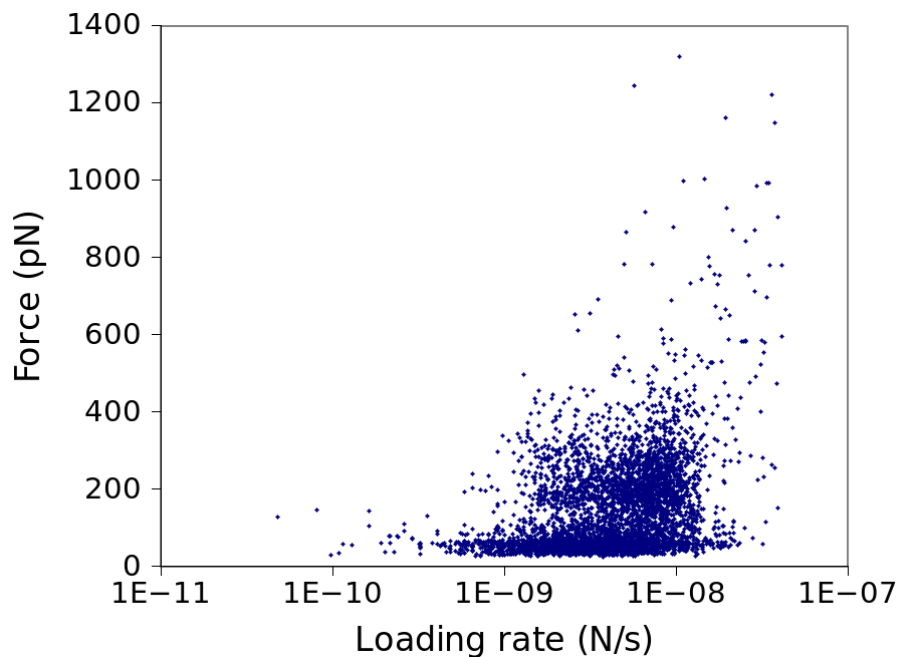
Figure 4.12: Scatter plot for the experimental results of the system HA New Caledonia *vs* 3'SLN.
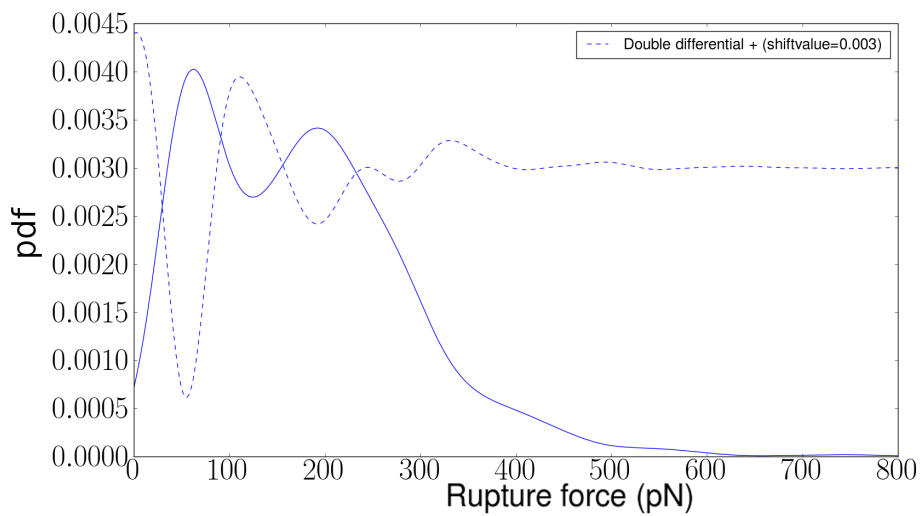
Figure 4.13: KDE of the points in the range of loading rate -19 to -18 (ln(N/s)) for the system HA New Caledonia *vs* 3'SLN.
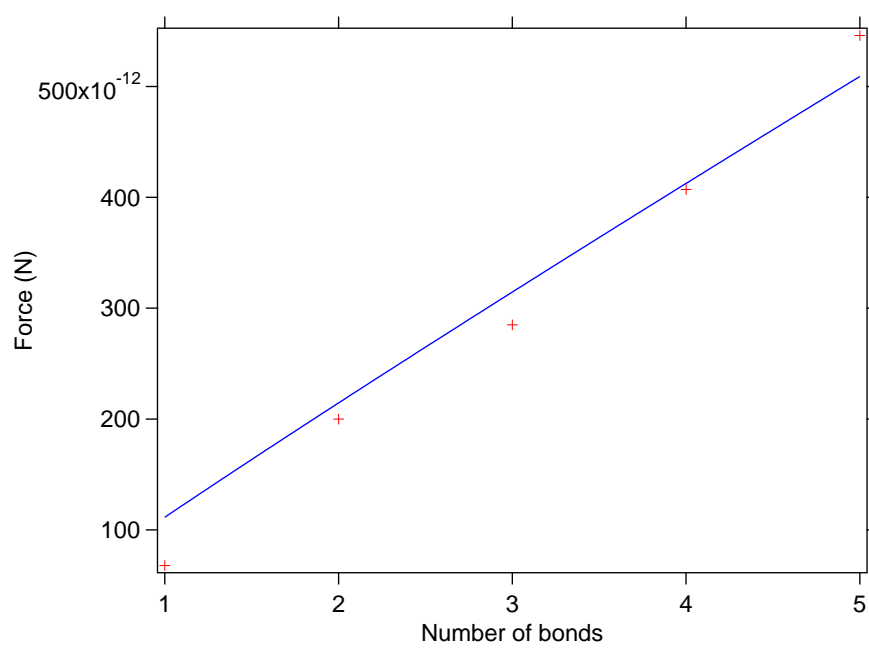
Figure 4.14: Fit of the maximum *vs* number of bonds for the system HA New Caledonia *vs* 3'SLN.

Chapter 5

An insight on the energy barrier

As we discuss in the "Introduction" the Bell-Evans model, which predicts the linear dependence of the most probable bond rupture force on the logarithm of force loading rate, is usually used to interpret the dynamic force spectroscopy experiment data. This model is consistent with the Kramers theory of the bond dissociation rate only if one assumes 1)an independence of the pre-exponential factors in the Kramers relation on the acting force and 2)a linear decrease of the dissociation barrier height on this same force. For this to be true a rather special shape of the interaction landscape is required.

In this chapter we are going to discuss a first order correction of the Bell-Evans model (first terms of corresponding Taylor expansions are taken into account), its implication for the interpretation of dynamic force spectroscopy experiment data and compare our model with the Monte Carlo simulation of a specially designed single molecule dynamic force spectroscopy experiment. In addition to the most probable bond rupture force, average rupture force values are also calculated. All approximations made and the range of applicability of the obtained results are carefully described and compared with those for some other models in the field.

5.1   First order approximation of Bell-Evans theory

We know that single molecule force spectroscopy is a versatile tool, which can often yield unique information concerning the energy landscape of individual ligand-receptor pairs; see e.g. [59, 60, 61, 62].

As we have discussed before the experiments consist in the stretching of the single molecule complex, whose two ends are fixed onto the AFM tip and appropriate substratum surfaces, with a force with a known time dependence,
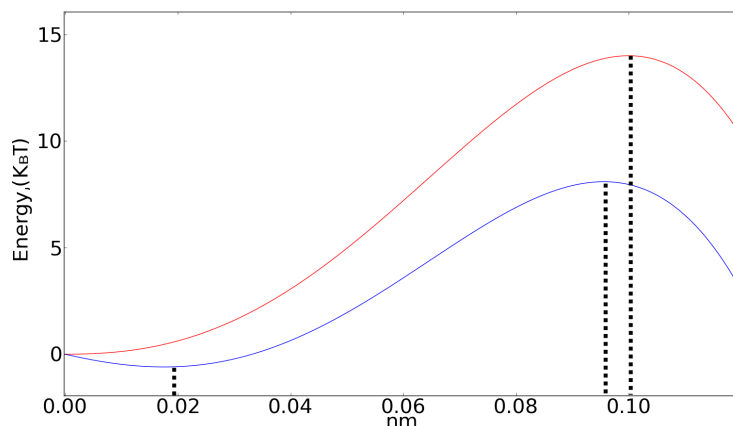
Figure 5.1: The red curve represent a possible energy landscape at zero force. The blue curve is the landscape when a force of approximately 250 pN is applied.

and measuring the dependence of the specific interaction rupture force F .

This is exactly the processing and interpretation of the data (either an average or the most probable rupture force can be used) which yields the informations about the interaction energy landscape: hence the details of its interpretation are very important.

From the very beginning of the force spectroscopy experiments, the Bell (also known as Bell-Evans) model for the dependence of the complex dissociation rate λ on the applied force [21, 22] has been used for the interpretation of experimental data, and up to now it remains the main approach in the field (although a number of refinements have been proposed during a few previous years[63]) The Bell model derives from the classical Arrhenius dissociation law with the assumed independence of the pre-exponential factor on the applied force and linear decrease of the free enthalpy on the same force. The transition rate obtained has been shown in the "Introduction".

The problem which immediately appears with this interpretation is that Bells model is only a zero-order approximation to the Kramers dissociation rate theory [20, 64]. The latter predicts that the dissociation rate is given by

$$\lambda = C\omega_{min}\omega_{max}exp(\Delta E/k_BT) \tag{5.1}$$

Here ΔE is the barrier height, that is the energy difference (for simplicity, from now on we will not distinguish energy, free energy or enthalpy and will always use the word "energy" as a generic term) between the maximum U_{max} and minimum U_{min} of the interaction potential U(x) which are attained respectively at the distances x_{max} and x_{min} (x is an appropriate reaction coordinate): $\Delta E = U_{max}(x_{max}) - U_{min}(x_{min})$, while $\omega_{max}, \omega_{min}$ are the oscillation frequencies corresponding to the potential minimum and maximum. We can take simply $\omega_{max}^2 = |\frac{d^2U}{dx^2}|_{x=x_{max}}$, $\omega_{min}^2 = |\frac{d^2U}{dx^2}|_{x=x_{min}}$ thus incorporating the

mass of interacting pocket and all other relevant constants like the diffusion co-
efficient, etc. into the constant C. When an external constant force F is applied
on the complex, the interaction energy changes from U(x) to U(x) - Fx (see
Fig.5.1 for an illustration) which leads to the change of parameters $\omega_{min}, \omega_{max}$
and ΔE but does not influence the constant C. Thus, instead of eq.2.17(in "In-
troduction") the dependence of the dissociation rate on the applied force in the
form $\lambda(F) = C\omega_{min}(F)\omega_{max}(F)exp(-\Delta E(F)/k_B T)$ should be used as a start-
ing point for the interpretation of the single molecule force spectroscopy data.
Unfortunately, concrete dependencies $\omega_{min}(F), \omega_{max}(F), \Delta E(F)$ are rather sen-
sitive to the details of the interaction potential which makes the search of such
a starting point a difficult task.

Recently, in a number of publications [63, 65, 66] the functions $\omega_{min}(F) =$
$a_1\varepsilon^{1/4}$, $\omega_{max}(F)a_2\varepsilon^{1/4}$, $\Delta E(F) = a_3\varepsilon^{3/2}$, where $\varepsilon = 1 - F/F_c$ and $a_{1,2,3}, F_c$
are appropriate constant parameters of the problem, were attempted, leading
to the dissociation rate law

$$\lambda(F) = C\varepsilon^{1/2}exp(-A\varepsilon^{3/2}) \tag{5.2}$$

(again, A, C are appropriate constants), and the subsequent use of this
law resulted in a rather different interpretation than the one given by simply
applying the Bells model on the force spectroscopy data. Essentially the authors
used the results of earlier works of Garg [67] and Kurkijärvi [68] where intrinsic
fluctuations in the superconducting ring closed with a Josephson junction are
studied and analytical results concerning different average values for such a
model are established. It was possible to borrow these results for discussion of
a new problem.

Such an approach definitely deepens our understanding of the dynamical
force spectroscopy but, in our opinion, it is still not fully satisfactory. The
main reason is that the use of the dissociation law in the form given by eq.
5.2, at least as this follows from the aforementioned papers, is justified only
if the most probable rupture force does not differ much from the critical force
F_c corresponding to the case when an applied force completely "smears out"
the dissociation barrier (potential $U(x) - Fx$ does not contain a maximum
any more). In many real experimental situations we are far from this, the most
probable rupture force is much smaller than the F_c value, so one needs a different
approximation going beyond that of Bell while being still consistent with the
Kramers dissociation theory. Such an approximation is derived here for both the
most probable and average rupture forces. The obtained equations are simple
and the limit of validity of the proposed approach is very clear.

Dissociation law

In the vicinity of the minimum or maximum of any interaction potential it
can be approximated by two first terms of the Taylor expansion: $U_{min} =$
$a_{min}x^2 + b_{min}x^3 + o(x^4)$ and $U_{max} = -a_{max}(x - x_0)^2 + b_{max}(x - x_0)^3 + o(x^4)$.
Here we put the minimum of the potential at point x=0 while its maximum
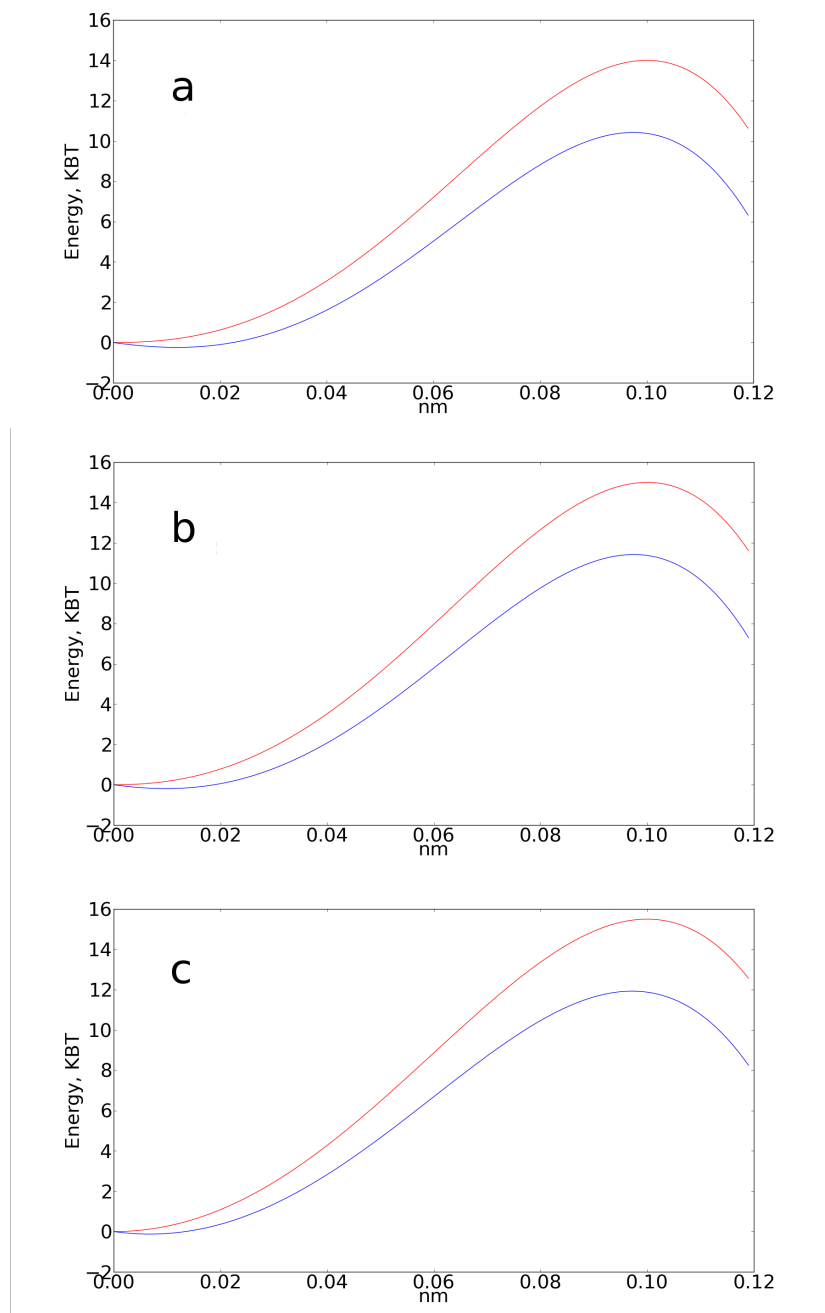is located at x_0; constants a_{min}, a_{max} are definitely positive while the signs of

Figure 5.2: In the figures we can see different potential at zero force (red line) for: a where the "B" value is 3.3, b where "B" value is 10 and c where the B value is 400. The blue lines are the correspondat potential where a force of 150 pN was applied.

b_{min}, b_{max} are arbitrary. If we apply the constant force F onto the system, potentials are deformed by $-Fx$ value and the positions of both minima and maxima shift on $\delta x_{min} = \frac{-a_{min} + \sqrt{a_{min}^2 + 3b_{min}F}}{3b_{min}}$, $\delta x_{max} = \frac{a_{max} - \sqrt{a_{max}^2 + 3b_{max}F}}{3b_{max}}$, which, for small enough forces, $F << \frac{a^2}{3b}$ is simply $\pm \frac{F}{2a_{min,max}}$. Thus, ignoring the quadratic and higher order terms on the force, the barrier height diminishes of the value $\Delta U = -Fx_0$ while the second derivatives of the potential at the positions of new maximum and minimum become $|\omega_{min}^2 = 2a_{min} + 3b_{min}\delta x_{min} = 2a_{min} + 3b_{min}F/a_{min}$ and, quite similarly, $|\omega_{max}^2| = 2a_{max} + 3b_{max}F/a_{max}$. Correspondingly, in the first approximation when $\beta F << 1$, for any one dimensional interaction potential we obtain the dissociation law following 5.1:

$$\lambda(F) = k_{\text{off}}(1 + \beta F)exp(Fx_0/k_B T) \tag{5.3}$$

where k_{off} and β are appropriate constants (k_{off} is, as usual, the dissociation rate at zero force conditions) and the parameter

$$\beta = \frac{3b_{min}}{4a_{min}^2} + \frac{3b_{max}}{4a_{max}^2} \tag{5.4}$$

may be negative or positive.

The most probable rupture force

Let us first calculate the most probable rupture force corresponding to the dissociation law 5.1. Our starting point is the expression

$$p(F)dF = \frac{\lambda(F)}{\dot{F}}exp\left(-\int_0^F \lambda(f)\frac{df}{\dot{f}}\right)dF \tag{5.5}$$

which describes the probability that the bond dissociates when the time-dependent applied force F(t) changes between the values of F and F+dF; here $\lambda(F)$ is the dissociation rate law and $\dot{F} = dF/dt$.

(Evidently, 5.5 is a product of the exponential factor which is the probability that the bond has survived during the force loading from zero to F and the factor $\lambda(F)dF$ that is dissociated when the force loading changes between F and F+dF; see [68] for general derivation). Expression 5.5 is valid for any arbitrary monotonic F(t). Now we will suppose that $\dot{F} = const$. For such a case the integration in 5.5 is elementary but the resulting expression is cumbersome and we will not present it here. The most probable force is given by the condition that derivative of 5.5 is equal to zero, which for this case gives:

$$\dot{F}\frac{d\lambda}{df} = \lambda^2(f) \tag{5.6}$$

In the introduction we have shown how it is possible to obtain the common used law of dissociation $F = F(\dot{F})$ from the rate $\lambda(F) = k_{\text{off}}exp(F\Delta x/k_B T)$. Let us use instead the dissociation law 5.3 to find the values of the most probable rupture force. Substituting 5.3 into 5.6 we get

$$\beta + \frac{1+\beta F}{k_B T}x_0 = \frac{k_{\text{off}}}{\dot{F}}(1+\beta F)^2 exp(Fx_0/k_B T) \tag{5.7}$$

From now on, let us introduce the dimensionless parameter $\tilde{B} = \frac{x_0}{\beta k_B T}$ which, again, may be positive or negative and define also the always positive parameter $b = |\tilde{B}|$. The first important peculiarity of 5.7 is that for negative values of β there exists some limiting value of the force F_0, such that if $F > F_0 = \frac{k_B T}{x_0}(B-1)$, this equation has no solutions due to its left part being negative, $\beta + \frac{1+\beta F}{k_B T}x_0 < 0$. For the case $B >> 1$ this is of no physical interest because for such large forces our approximation in any case becomes inapplicable (see below). If $B < 1$, this equation has no solution for any applied force, but, again this is of no physical interest, because for any sign of $\tilde{B}$, in the case $B < 1$ the presented approach may be applied only in a very narrow range of forces (see below) and will not be considered here. After taking the logarithm and rearranging 5.7 we obtain eq. 5.8.

$$ln\left(\frac{x_0}{k_B T} + \beta + \frac{\beta F}{k_B T}x_0\right) = ln\left(\frac{k_{\text{off}}}{\dot{F}}\right) + 2ln\left(1+\beta F\right) + \frac{Fx_0}{k_B T} \tag{5.8}$$

First, we would like to note that this equation as well as eq. 5.7 can be easily rewritten as an explicit dependence $\dot{F} = \dot{F}(F, \beta, x_0, k_{\text{off}})$, where the most probable rupture force F is an argument and $\beta, x_0, k_{\text{off}}$ are parameters of the complex under the study, and then the inverse function $F = F(\dot{F}, \beta, x_0, k_{\text{off}})$ can be plotted and used for comparison with the experiment without further additional simplifications presented below. In the range of the validity of our approximation, when $\beta F << 1$ (see below what this condition physically means), we have that $2ln(1+\beta F) \cong 2\beta F$ and that the term $\frac{\beta F}{k_B T}x_0$ in the l.h.s. of 5.8 is a small correction to $\frac{x_0}{k_B T}$, thus 5.8 is reasonably approximated eq.5.9.

$$ln(\frac{x_0}{k_B T} + \beta) = ln(\frac{k_{\text{off}}}{\dot{F}}) + 2\beta F + \frac{Fx_0}{k_B T} \tag{5.9}$$

If $|\beta|$ is much smaller than $x_0/k_B T$ (large B case) we have in the first order of the Taylor expansion eq. 5.10.

$$F = \frac{k_b T}{x_0}(1 - \frac{\beta k_B T}{x_0})ln(\frac{\dot{F}x_0}{k_{\text{off}}k_B T} + \frac{\beta k_B T}{x_0}) \tag{5.10}$$

Eq.5.10 can be rewritten as 5.11.

$$F = \frac{k_B T}{x_0}(1 - \frac{1}{\tilde{B}})ln\dot{F} + \frac{k_B T}{x_0}(ln\frac{x_0}{k_{\text{off}}k_B T} + \frac{1}{\tilde{B}}) \tag{5.11}$$

This gives the required correction to the Bell dissociation law; clearly 5.10 coincides with eq. 2.17 (in "Introduction") when $\beta = 0$. Of course, as usual in single molecule force spectroscopy, it should be that $\frac{\dot{F}x_0}{k_{\text{off}}k_B T} > 1$, i.e. that $\dot{F} > \frac{k_{\text{off}}k_B T}{x_0}$ (negative values of the force have no sense). Physically it means

that the force loading should be fast enough to overcome the dissociation dynamics dominated simply by k_{off} when the measured "most probable rupture force" is around $\dot{F}/k_{\text{off}}$, just the value attained at a moment of "natural" thermal fluctuations-induced force-free bond breaking. What the condition $\beta F << 1$ actually means? It means simply $F << B\frac{k_B T}{x_0}$, which for $B \cong 10$, say, effective barrier thickness x_0 around 0.1 nm (which was often reported in force spectroscopy experiments) [59, 60, 61, 62, 57, 69], and the room temperature case $k_B T \cong 4 \cdot 10^{-21}$ J, gives that the most probable rupture force should be smaller than ca. 400 pN which is not so restrictive. At the same time, only very small forces are admissible for the inverse case $B << 1$ and this is the reason why we do not analyze it here.

The average rupture force

In many cases the use of average rupture force value rather than the most probable rupture force value might be useful. In fact, when the loading rate condition are not to low (and this happen in the majority of the experiments), these two values are in good agreement. Moreover the average force can be obtained rapidly from the data, when instead, the most probable require a more accurate procedure.

For this reason here we present the corresponding derivation pertinent to the dissociation law 5.3. This same law is not exactly of Gargs form $\lambda(\varepsilon) = A\varepsilon^{a+b-1}exp(-B\varepsilon^b)$ and one needs some efforts to adopt his approach, especially for the case of positive β. The corresponding consideration (whose details and all the approximations used are given in the Appendix) brings the following results: for negative β with the precision up to $ln|lnX|$ we have for $\varepsilon = 1 + \beta F$: $< \varepsilon >= (X + lnX + \gamma)/B$, $< \varepsilon^2 >=< \varepsilon >^2 +\frac{\pi^2}{6B^2}$, where $\gamma = 0.5772...$ is the Euler-Mascheroni constant, $X = ln(\frac{A}{|\beta|\dot{F}B^2} = ln\frac{k_{\text{off}}k_B T}{x_0 B\dot{F}} + B$, $B = \frac{x_0}{|\beta|k_B T}$, $A = k_{\text{off}}exp(B)$. For positive β we have $< \varepsilon >= (-X - ln(-X) - \gamma)/B$, $< \varepsilon^2 >=< \varepsilon >^2 +\frac{\pi^2}{6B^2}$. In terms of the force the results read:

$$< F >= \frac{k_B T}{x_0}ln(\dot{F})+\frac{k_B T}{x_0}(ln\frac{x_0}{k_{\text{off}}k_B T}+ln(ln\frac{k_{off}k_B T}{B\dot{F}x_0}+B)-lnB-\gamma) \quad (5.12)$$

for negative β,

$$< F >= \frac{k_B T}{x_0}ln\dot{F}+\frac{k_B T}{x_0}(ln\frac{x_0}{k_{off}k_B T}+ln(-ln\frac{k_{off}k_B T}{B\dot{F}x_0}+B)-lnB-\gamma) \quad (5.13)$$

for positive β, and

$$\langle F^2 \rangle - \langle F \rangle^2 = \frac{\pi^2}{6}\left(\frac{k_B T}{x_0}\right) \quad (5.14)$$

for both cases. Eq. 5.14 also shows that, in contrast with the simplest Bell case, the difference between $< F^2 >$ and $< F >^2$ is now present. If, what is the

most typical case, Taylor expansion $ln(B \pm ln\frac{k_{off}k_BT}{B\dot{F}x_0} \cong lnB \pm \frac{1}{B}ln\frac{k_{off}k_BT}{B\dot{F}x_0}$ can be used, we get

$$<F> = \frac{k_BT}{x_0}(1+\frac{1}{B})ln\dot{F} + \frac{k_BT}{x_0}(1+\frac{1}{B})ln\frac{x_0}{k_{off}k_BT} + \frac{k_BT}{x_0}(-\gamma - \frac{lnB}{B}) \quad (5.15)$$

for negative β, and

$$<F> = \frac{k_BT}{x_0}(1-\frac{1}{B})ln\dot{F} + \frac{k_BT}{x_0}(1-\frac{1}{B})ln\frac{x_0}{k_{off}k_BT} + \frac{k_BT}{x_0}(-\gamma + \frac{lnB}{B}) \quad (5.16)$$

for positive β cases. These are almost the same as the expressions for the most probable rupture forces given by eq. 5.10.

When can it be applied?

We would like to end this section with a short discussion on the usage of a dissociation law in the form $\lambda(F) = C\varepsilon^{1/2}exp(-A\varepsilon^{3/2})$ to describe the dynamic force spectroscopy data. Suppose that the interaction potential is such that a single Taylor expansion $U = ax^2 - bx^3$ can be used to describe both the minimum and maximum of the interaction energy landscape (hence the term $-bx^3$ is certainly negative as written). Then the positions of maximum and minimum when the force is applied are $x_{min,max} = \frac{a \pm \sqrt{a^2 - 3bF}}{3b} = \frac{a \pm \sqrt{3b\Delta F}}{3b}$, where $\Delta F = F_c - F$, $F_c = \frac{a^2}{3b}$. Substituting this into $U = ax^2 - bx^3 - Fx$ we see that the barrier height when the force F is applied is given by $U_{max}(x_{max}) - U_{min}(x_{min}) = \frac{4(3b\Delta F)^{3/2}}{27b^2}$ and that $\omega_{max}^2 = \omega_{min}^2 = \sqrt{12b}\Delta F^{1/2}$, hence the dissociation law 5.2 follows.

What B corresponds to this case? One has $x_0 = \frac{2a}{3b}$, $\Delta E_0 = \frac{4a^3}{27b^2}$, so that $\beta = -\frac{3b}{2a^2} = -\frac{x_0}{3\Delta E_0}$: for the case at hand β is negative and $B = \frac{x_0}{|\beta|k_BT} >> 1$ whenever $\Delta E_0 >> 3k_BT$. The latter always holds for real experimental situation. From this consideration, we easily see how restrictive the conditions of the validity of the dissociation law 5.2 are. It can be applied reasonably only when the acting force is really close to the critical force F_c for which the interaction well disappears completely. Therefore it makes no sense to use the conclusions obtained when using 5.2 to discuss the force spectroscopy data pertinent for small forces, and it is completely meaningless to attempt such a use to interpret the value of k_{off}, [63], which is a limit for zero applied force.

Application of the results to analyze dynamic force spectroscopy data

In this Section we consider how the presented approach can be used to analyze the dynamic force spectroscopy data; viz. we first attempted some refinement of the interpretation of real experimental data pertinent to avidin−biotin and streptavidin−biotin pairs, and then also compare our analytical formulae with the results of Monte Carlo simulation of specially designed dynamic force spectroscopy experiment.

For such a consideration it seems quite natural to select the (strept)avidin−biotin system which is one of the most profoundly studied in the field, despite the circumstance that even for this protein pair a number of contradictory results persist, see e.g. [70] for a discussion. Having in mind that our results are best applicable to relatively small forces and small force loading ranges, it looks natural to use the results presented by Merkel et al. [71] and de Obrowaz Piramowicz et al. [57]. Unfortunately, the estimated barrier width $x_0 \cong 0.5$ nm given for small force loading rates in [71] makes the range of applicability of our approach too narrow (see above) hence we will concentrate only on the latter paper where the estimated barrier widths, force loading rates used and the smallest measured specific interaction forces apparently enable us to apply the proposed method.

For the force loading rates in the range of ca. 300-1700 pN/s, de Obrowaz Piramowicz et al. [57] reported different dissociation barrier parameters for avidin−biotin and streptavidin−biotin (simple Bell model was used in both cases to process the data): respectively $x_0 = 0.07$ nm k_{off} 0.25 s^{-1} and $x_0 = 0.08$ nm, $k_{off} = 0.56$ s^{-1} [57]. At the same time, for both protein pairs they obtained quite similar results for larger force loading rates, ca. 1700−9600 pN/s, and these observations are used by the authors to discuss the difference in outer activation barriers for these pairs equal to 0.86 kT simultaneously with the equivalence of their inner activation barriers. However, it is instructive to see if it is possible or not to associate the observed difference with the different shapes of corresponding interaction landscape (i.e. different B or β in our terminology) rather than with the difference in barrier heights and width. Let us (arbitrary) put $\beta = 0$ for avidin−biotin energy landscape. Then from the formula 5.11 it is quite easy to see that a rather moderate $\tilde{B} = \frac{x_0}{\beta k_B T} = -7$ value may quite well explain the larger barrier width measured for the streptavidin−biotin pair. Other way around, putting $\beta = 0$ for streptavidin−biotin energy landscape one needs to use $\tilde{B} = 8$ for avidin−biotin to obtain the same value of the barrier width. To illustrate our approach further, in Fig.5.3 we show how well the green straight line $B = \infty$ corresponding to the Bell model-based interpretation of the avidin−biotin dynamic force spectroscopy data (parameters of Ref. [57]) can be approximated varying B, x_0, k_{off} parameters of our model. Taking "by hand" for the avidin−biotin protein pair at question the value $k_{off} = 0.56$ s^{-1}, that is the value reported in [57] as characterizing the dissociation of strepavidin−biotin pair at low force loading rates (remember that exactly this same value of k_{off} is a starting point for the determination of the barrier height), we obtain a sufficiently reasonable correspondence for $\tilde{B} = 5$ and $x_0 = 0.04$ nm (dashed line). For a typical precision of real single molecule dynamic force spectroscopy experiment (of which the aforementioned de Obrowaz Piramowicz et al. work [57] is fairly good example), this would be not at all easy to select which curve is to be used for an approximation of an actual experimental data, and, correspondingly, to extract the values of the barrier width and height.

By no means above example should be understood as a criticism of an interesting research presented in [57]. In our opinion, this situation is quite common in the force spectroscopy field: always neglected correction factor consid-
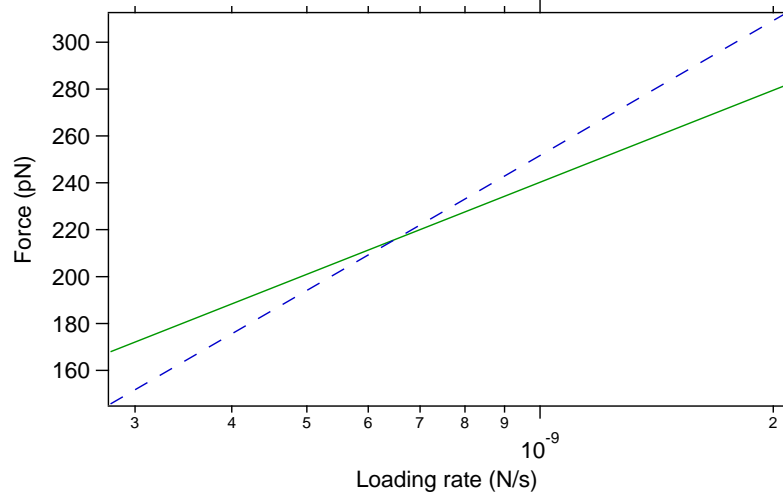
Figure 5.3: Showing an approximation of the dynamic force spectroscopy data for avidin−bioting pair. Green curve: $x_0 = 0.07$ nm, $k_{off} = 0.25$ s^{-1}, $B = \infty$ (Bell model, parameters taken from Ref. [57]). Blue curve: equation 5.11 with $x_0 = 0.04$ nm, $k_{off} = 0.56$ s^{-1}, $\tilde{B} = 5$.

ered in the paper, coefficient β (or B), does matter in the interpretation of the force spectroscopy data, especially in the region of relatively small loading rates, and it should be taken into account to obtain correct interaction landscape parameters. Very little is known about this coefficient (energy landscape shape beyond the simplest parabolic approximation) up to now, but still we believe the following example, where we borrow the data from the Table 1 of the paper [70], will be interesting to demonstrate how important the term β might be. In this paper the authors, using essentially the molecular dynamics calculations of Merkel et al. [71] and Izrailev et al. [56], give the values of energies and curvatures at corresponding minimums and maximums for three neighboring wells and barriers of the interaction energy landscape pertaining for streptavidin−biotin interaction. Applying this data and requiring that the Taylor expansion $U_{max} = -a_{max}(x - x_0)^2 + b_{max}(x - x_0)^3$ of the interaction potential in the vicinity of the first interaction potential maximum (located at $x_0 = 0.31$ nm with the energy 32 $k_B T$ and the curvature $2a_{max} = 750$ $k_B T/nm^2$ [70]) gives also the position of a second interaction potential minimum (located at $x_0 = 0.39$ nm and with the energy 5 $k_B T$), we get an estimation and thus $\beta = \frac{3b_{max}}{4a_{max}^2} \cong -0.3$ $nm/k_B T$. No data for the deepest energy well and hence for the term $\frac{3b_{min}}{4a_{min}^2}$ in expression 5.4 is available, and this is an additional complication, but this example apparently demonstrates that the value of the parameter β (negative in the case) may be quite comparable with the $x_0/k_B T$ (equal to $\cong 0.3$ $nm/k_B T$ for the same system).

Monte Carlo simulation of dynamic force spectroscopy experiment and its comparison with the model

Additionally, we test our model using the Monte Carlo simulation of the dynamic force spectroscopy data. To describe the interaction potential, the simplest nontrivial case of a trinomial $U = ax^2 + bx^3 + cx^4$ having a maximum at the point $x_0 = 0.1$ nm and barrier height (at zero force) of $\Delta U = 13 - 15\ k_BT$ was used. For the potential given above it turned out quite easy to simulate any value of B, including even very small $B << 1$, for positive β (taking positive a and b and negative c coefficients) while only sufficiently large values of B, $B \geq 35$, can be simulated for negative β. More complex polynomials modeling the case of relatively small B for negative β are needed; this situation will be considered elsewhere. When the force F is applied, the potential evidently takes the form $\tilde{U} = ax^2 + bx^3 + cx^4 - Fx$. The position of minimum and maximum x_{min}, x_{max} for this potential were determined numerically by solving a cubic equation $2ax + 3bx^2 + 4cx^3 - F = 0$ (Cardano formula was applied), and then the energy barrier $\Delta U = \tilde{U}(x_{min}) - \tilde{U}(x_{max})$ as well as the values of $\omega_{min}^2 = 2a + 6bx_{min} + 12cx_{min}^2$, $\omega_{max}^2 = -(2a + 6bx_{max} + 12cx_{max}^2)$ were calculated.

These values were utilized to calculate the force-dependent dissociation rate $\lambda(F) = C\omega_{min}(F)\omega_{max}(F)exp(-\Delta U(F)/k_BT)$. Constant "C" was selected to get the desirable value of $k_{off} = \lambda(0)$; for our modeling we use $k_{off} = 0.01$ s^{-1}. The procedure exploited for Monte Carlo simulation of the dynamic force spectroscopy experiment using the so obtained parameters of the interaction potential was similar to the protocol reported in literature [8, 41]. Earlier this same procedure has been tested in our work devoted to the prediction of the unfolding forces of multi-modular proteins[72]. For simulations, the pulling speed in the range of 0.02 to 10 nm/s and a spring constant of the cantilever of 0.05 N/m were used. Worm-like chain dependence of the end-to-end protein complex distance on the applied force ([73]) with the total contour length of 27.5 nm, persistence length of 0.3 nm and room temperature (300 K) was assumed. The time step of the simulation procedure was set in such a way that for each increase of the complex end-to-end distance equal to 5 pm (at any loading rate) we control the probability to dissociate. In Fig.5.4 we present a data of the Monte Carlo simulation of the most probable rupture force together with the theoretical curve given by eq.5.11 of our model. We have to stress that the "C" factor of the eq.5.1 used for the generation of the data was set in such way that for every simulation with a different "B" we have $\lambda(0) = 0.01$, this is equivalent to changing the viscous drag in the Bell-Evans theory.

This Figure well illustrates the applicability of our model (as corresponds to its validity range, for $\tilde{B} = 10$ case the model ceases to work properly for forces larger than ca. 300 pN), and also gives an idea where and when the very concept of the most probable rupture force is appropriate: for the lowest force loading rates presented here, this concept does not work because the force loading is so slow that dynamics given by the factor $\dot{F}/k_{off}$, see above, dominates. The data obtained for the average data forces is, not surprisingly, quite similar (compare

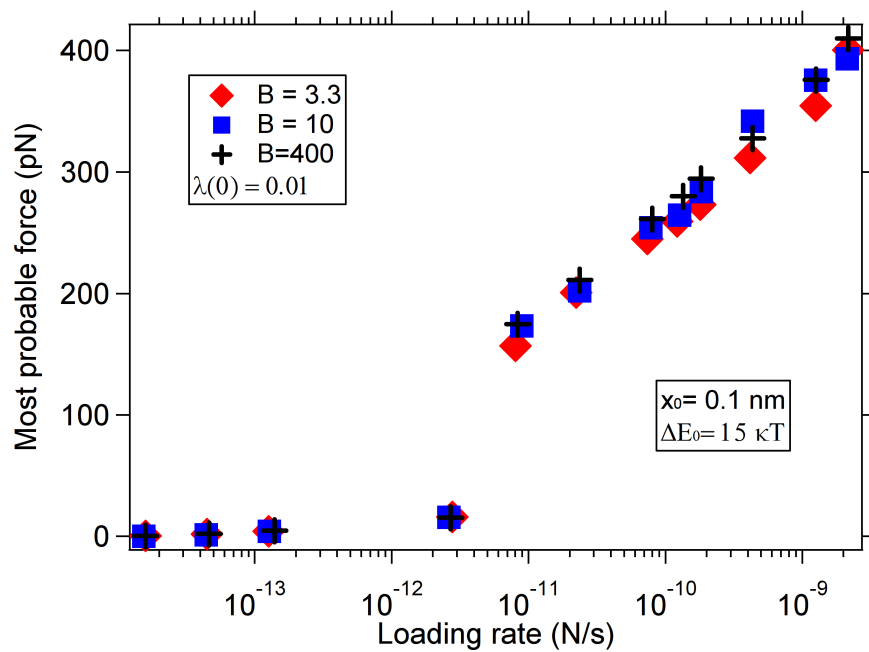Figure 5.4: Here we have plot the most probable rupture force for three different simulations. The parameter "B" of the barriers were respectively: 3.3 for the red diamonds, 10 for the blue square and 400 for the black cross. The similarity of the results can be abscribed to the fact that the "C" value of the equation 5.1 was selected in such way that the $\lambda(0)$ was the same for all the simulations.

equations 5.11 and 5.12 5.13) and by this reason not presented here. We would like only to note that systematically smaller rupture forces (of around 30-50 pN on average) when compared with the most probable rupture forces (compare equations 5.11 and 5.12 5.13 again) were observed in simulations.

Conclusion

In this section we derived a refinement to the well known Bell-Evans law of the dependence of the most probable bond rupture force F as a function of the force loading rate . This refinement could be called "a first order approximation $F = F(ln\dot{F})$ law" in the following sense. Kramers' dissociation theory predicts that the bond dissociation rate is given by an expression $\lambda(F) = C\omega_{min}(F)\omega_{max}(F)exp(-\Delta E(F)/k_BT)$, where all the parameters $\omega_{min}(F)$, $\omega_{max}(F)$, $\Delta E(F)$ depend on the applied force. If one ignores the force dependence of the pre-factors $\omega_{min,max}(F)$ and puts a linear dependence $\Delta E(F) = \Delta E_0 - Fz$, they obtain the famous Bell-Evans $F = F(ln\dot{F})$ law which, according to the aforementioned derivation procedure, can be called "a zero order approximation law": it has the serious advantage of being very simple and clear, but evidently it is not always sufficient. The natural next approximation is to take the first Taylor-expansion terms for both $\omega_{min,max}(F)$ and $\Delta E(F)$ dependencies and, using corresponding dissociation law as a starting point, to derive "a first order approximation $F = F(ln\dot{F})$ law", and this is exactly what is has been presented in this chapter. We also analyze the applicability range of the obtained results and show that albeit this range is not very broad, there is definitely enough room for the application of our results in practice. An example given in Section 3 shows that for quite ordinary protein pairs the parameter β may be well comparable with the x_0/k_BT. If so, the effect considered in the paper is of utmost importance for dynamic force spectroscopy data interpretation. In any case, all corresponding aspects deserve a deeper study and further investigations, both of them currently under way.

Chapter 6

Laminin binding protein pulling

6.1   The laminin binding protein

One of the subjects of this thesis was the study of the ligand receptor interaction. This mechanism, of which we spoke in the "Introduction", is present also when a virus attaches to the cell surface to start the infection process. The Laminin binding protein (LBP) [74] is a protein that mediates the interaction with laminin, a glycoprotein that resides on the cell surface. It is known that the LPB is used by many viruses as the first step of the infection process. The initial aim of our experiments was to deepen the knowledge of the flaviviral penetration into the cell membrane by force spectroscopy measurements, and to understand if this interaction changes when modifying the pH. For such purpose we used the LBP and the protein "E11". E11 is is a recombinant protein of around 200 amminoacid length, which contains a sequence responsible for the specific interaction with LBP. Preliminary results has been obtained on this topic and they have been used as starting point for further studies in Russia, however their scope goes beyond this thesis.

Moreover it has been observed that the expression of LPB increases considerably in the metastatic tissues and it is also known that the pH decreases in metastatic tissues [75]. However, data coming from NMR [76] showed that the pH change is due to the extracellular matrix environment that becomes more acid. The change in the behavior of the LPB and the decrease of the pH in metastasis could be then correlated.

Here we present the results obtained by single molecule force spectroscopy of this protein/ligand complex. Unexpected results arose from our experiments at low pH.
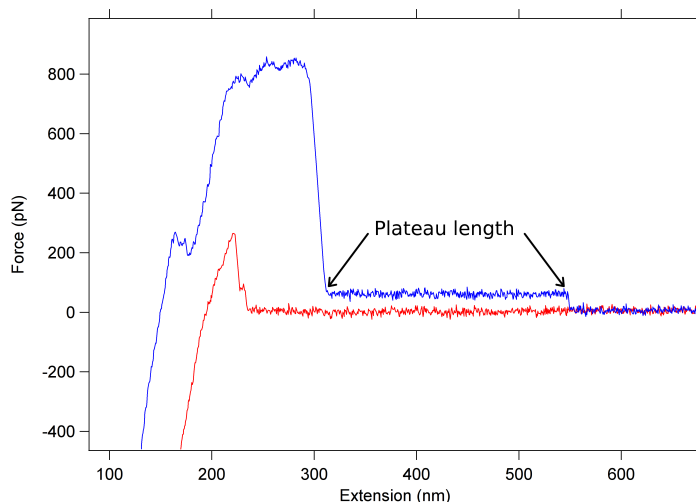
Figure 6.1: In the image we can see the typical force curve obtained with the LBP with the disodium phosphate buffer.

6.1.1   Experiments and results

Originally the experiments between E11 and LBP where performed as described in "Introduction", attaching covalently E11 on AFM tip and LPB on the surface with the APTES/glutaraldehyde procedure. The data obtained with this methodology were interesting and these experiment have been continued by B. N. Zaitsev [1], however they go beyond the scope of this thesis.

Experiments at low pH were performed in disodium phosphate buffer ($Na_2HPO_4 \cdot 12H_2O$) 50 mM plus citric acid 25 mM. 50 ml of $Na_2HPO_4 \cdot 12H_2O$ at 0.2 M was mixed with 50 ml of citric acid 0.1 M, then 100 ml of distilled water was added. The final pH of the buffer was 5.11.

One of the most interesting result obtained with the LBP protein regarded its behavior at low pH, where almost all the curves present a plateau signal. Control experiments made with LBP at pH 7 proved that the plateau effect could be abscribed only to the LBP.

In the Fig.6.1 we can see that the LBP gives rise to a region at constant force. This behavior, at the beginning unexpected, can be explained with an unzipping process of the LBP. At the moment however, there are no structural data that can tell us from which configuration this unzipping occur. We could hypothesize that the LPB lose its structure and create a series of bonds with the substrate.

This change in the behavior of this protein could correspond to a phase transition. In fact, experiments performed at pH 7 did not result in any "plateau"

---

[1]State Research Center of Virology and Biotechnology "Vector", Ministry of public health and social development, Novosibirsk, Russia

Figure 6.2: In the top graph "1" we can see the KDE for the experiment A (429 events). On the graph "2" it is possible to see the deconvolution obtained.

effect.

We have analyzed the distribution of lengths for the constant force transition. We measured the plateau length starting right after the drop in the force due to the aspecific interaction. Several curves presented also a different pattern where we had multiple plateaux in series, each at a different force. The two experiments performed contain, respectively, 429 measures (experiment A) and 1322 measures (experiment B).

We performed a KDE to recover the distribution of plateau lengths, and then we have analyzed them using Igor with the multipeak analysis. This last tool allows to deconvolve the KDE with sub-gaussian probability distributions.

The two experiments performed showed a really good agreement on a plateau length value corresponding to the fully unfolded length of the molecule. In fact, in both experiments we get a value close to 100 nm. Considering that the LPB is composed by 295 ammino acids and that the length for one of them is between 0.3 and 0.35 nm we can estimate a fully unfolded protein length between 88 nm and 103 nm.

However, looking at the following peak maximum the situation is more complex. For the experiment "A" we can say that the peaks at higher lengths correspond to the dimers and trimers of the LBP, in fact the length seem to be, respectively, two and three times the length of the fully extended protein.

Instead for the experiment "B" we obtained two peaks which are difficult to interpret, one situated at 174 nm and the other at 221 nm. Both these peaks are close to 200 nm, so it is possible that both correspond to dimers. The difference in the two experiments could possibly be ascribed to the different deposition times, that were not completely controlled.
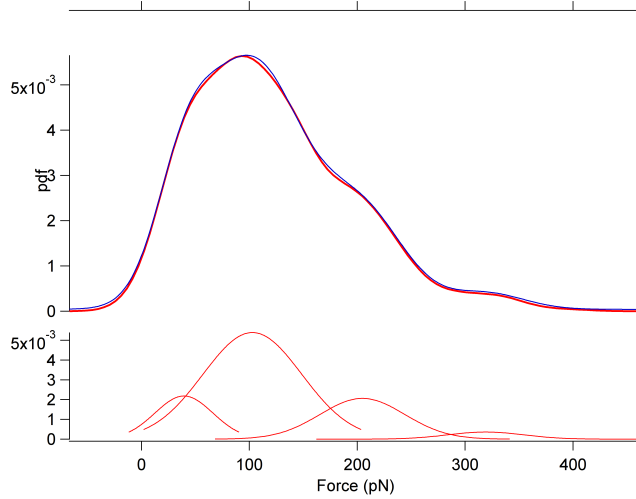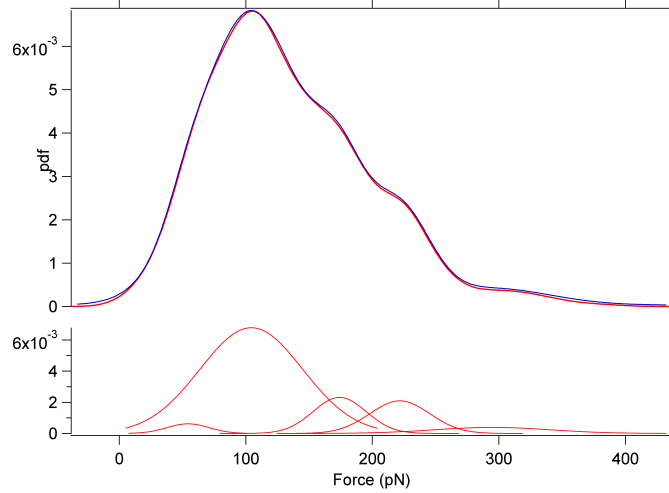
Figure 6.3: In the top graph "1" we can see the KDE for the experiment B (1322 events). On the graph "2" it is possible to see the deconvolution obtained.

Probability peaks characteristics					
Type	Location	Local Sigma	Amplitude	Area	FWHM
Gauss	39.4	0.6	0.0021	0.14	3.1
Gauss	102.8	1.0	0.0053	0.62	108.2
Gauss	204.8	1.8	0.0020	0.20	92.5
Gauss	319.2	2.2	0.00035	0.03	83.1

Table 6.1: Paramters of the fit obtained with Igor for the experiment A (429 events).

6.1.2   Conclusion on LBP

The experiments at pH 5 demonstrate a change in the behavior of the LBP. The force plateau is compatible with an unzipping mechanism (a similar behavior has been observed in amyloids proteins by Karsai et Al.[77] where they were unzipping $\beta-$sheets from amyloid proteins observing plateau). The deconvolution of the lengths' distributions estimated by KDE demonstrate how the prominent peak is in good agreement with the full contour length of the LBP. This fact suggests that the plateau is indeed due to the protein undergoing a phase transition. Other peaks appear in the distributions, they may be due to the di- and tri- merization of the LBP.

Shorter peaks appear too in the distribution: this fact is also expected, because we cannot expect to pick the LBP always at the very end with our tip.

Probability peaks characteristics					
Type	Location	Local Sigma	Amplitude	Area	FWHM
Gauss	54.3	0.41	0.00061	0.025772	39.1
Gauss	104.3	0.24	0.00678	0.69	95.7
Gauss	174.1	0.57	0.00231	0.12	49.3
Gauss	221.6	0.76	0.00208	0.12	55.9
Gauss	294.7	3.52	0.000385	0.044	108.2

Table 6.2: Parameters of the fit obtained with Igor for the experiment B (1322 events).

Chapter 7

Nano construct for protein-protein interaction

In this chapter we will discuss the development of a new methodology to study "lock and key" interactions and, on the other side, a possible strategy to make a material that assembles from the constituents with a particular designed structure.

For this purpose we used a multimodular I27 protein and another multimodular construct composed by staphylococcal nuclease and I27. These molecules are used as "handles" to study the molecules that form the "lock and key" bond. The general idea is to have two different constructs, an "Handle-Lock" and an "Handle-key" that can then bind and assemble in a "Handle-lock-key-Handle" construct.

To obtain such a construct the plasmid coding for the handles has been concatenated with the plasmid of the lock/key molecules and then the proteins have been expressed (separately) in E.coli.

This strategy and the corresponding detailed experimental procedures have been discussed in the work of Kim et al. [78] [79] (see their supporting material). The present chapter will be more focused on the analysis of the data obtained.

A similar technology has been used before to study the properties of the $\alpha-$synuclein inserted in a multimodular construct [80]. The idea of use the handles is similar, nevertheless the interactions that we are going to study present some complication that we are going to discuss.

7.1 Streptavidin tetramers

We start our study with the investigation of the streptavidin tetramer. For this purpose we used two different constructs composed by:

- I27 and a streptavidin monomer,

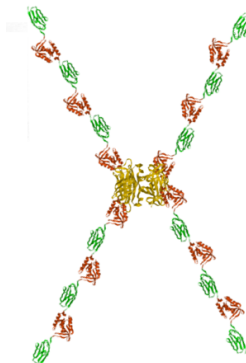- I27, sthapylococcal nuclease (SNase) and a streptavidin monomer (SM).

Figure 7.1: In the figure we can see a representation of the $[(I27 - SNase)_3SM]_4$ tetramer. If we pull such construct (from just two ends) with the AFM, to be sure that we are pulling also the streptavidin, we need to see at least 4 peaks of SNase. I27 is represented in green, Snase in red.

The SNase have been studied by Wang et al. [81] and its SMFS force curve pattern is characterized by peaks of 46 nm of contour length at low unfolding force (around 25 pN at 500 nm/s of pulling speed).

The SM has the capacity to form tetramers[82] and this property could be exploited to create biologically inspired materials and structures. In particular this protein, if bound with a determined chain, could be used as "hub" for a material. The basic idea is in fact to append a construct of a known nature to the SM and allow it to self-assemble. We however must remember that not any kind of polymer could be attached, because, depending on the attached moiety, the capacity of SM to self-assemble could be seriously inhibited.

There are severals papers that studied the bond between the biotin and the streptavidin [57] [70] [83]. Most of these studies used a direct interaction between these two molecules, attaching one of them on the cantilever tip and the other on a flat surface. Intriguingly, there are no studies that investigate the properties of the streptavidin tetramers by AFM.

The construct used were then, respectively $(I27_6SM)_4$ and$[(I27\text{-}SNase)_3SM]_4$. Both these constructs give us a reference pattern, where the equally spaced peaks (in terms of WLC/FJC-fitted contour length) allow us to correctly classify the interaction and be sure that only a single molecule is pulled.

In particular we know that I27 modules give a sawtooth pattern of around 200 pN force at 500 nm/s of pulling speed, when SNase, at the same conditions, yields peaks of at most 20-30 pN. These constructs give us some advantages in force spectroscopy. In the case of the $(I27_6SM)_4$, if we see at least seven unfolding force peaks, we are sure that also the SM tetramer have been pulled. For the case of $[(I27 - SNase)_3SM]4$, due to the different nature of the SNase and I27 we are sure that the SM tetramer is pulled when we see four SNase peaks.

Figure 7.2: In the figure we can see the comparison between the KDE of the construct with $I27_7$ (violet line) with the construct $I27_6SM$ (blue line). The KDE was obtained with a uniform kernel [38].

However, another complication arises from this approach. The rupture of the avidin tetramer does not give a reference pattern. In fact, if we look at a force extension curve, when there is the rupture of the avidin tetramer or when a detachment happen we have exactly the same effect that consist in the relaxation of the cantilever with no other successive events. The chain, in both case, do not connect anymore the cantilever to the surface.

Then, to correctly characterize the behavior of the SM, it is mandatory to compare the statistic of the "rupture force" [1], with the statistic of the detachment in the "non-SM" construct.

For this purpose two kinds of experiments has been performed:

- collecting the rupture forces in the SM construct,

- collecting the statistic of the detachment forces of the correspondent "non-SM" construct.

In Fig.7.2 we can see the comparison of the PDF (obtained with the KDE method) between $I27_7$ and $I27_6SM$. It is possible to notice that the PDF for the $I27_7$ has a long tail that reaches 700 pN, while the one for the the construct $I27_6SM$ ends at 500 pN and has an hump at 400 pN. It is arguable that the difference in the tails is due to the presence of the SM that stops the pulling before the forces higher than 500 pN are reached, moreover the hump at 400 pN could be also an effect of the SM. However, a basic objection can be done to this

[1]We define "rupture force" as the characteristic force for the last event of a force curve before return to the baseline in the construct that contain SM. The group of this kind of events, in the construct that contain SM, contains the rupture of the avidin tetramer and also the classical detachment of the polymer from the tip or the surface.
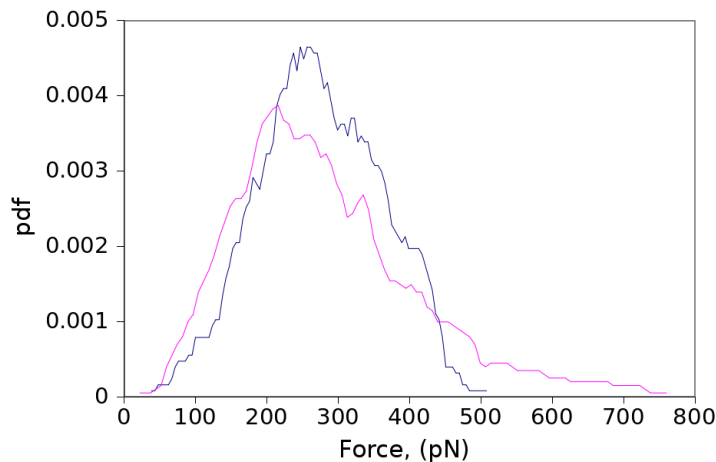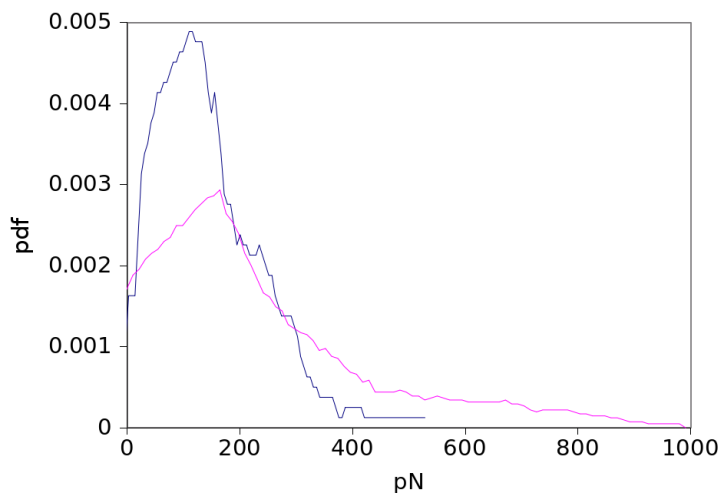
Figure 7.3: In the figure we can see the comparison between the KDE of the construct with $I27 - (Snase - I27)_3$ (violet line) with the construct $(I27 - SNase)_3SM$ (blue line). The KDE was obtained with a uniform kernel [38].

method. Using I27 modules, we need to see their peaks to be sure we pull the construct. The peaks, in this construct, appear at 200 pN and this means that all the rupture forces that appear below 200 pN are ignored. Substantially this particular system behave like a high pass filter. It is then clear that, to improve the analysis of the rupture forces, a construct like $(I27 - SNase)_3SM$ should be used. As aforementioned, at 500 nm/s, the SNase unfolds at 25 pN and, to be sure that we are also pulling the SM tetramers, with this construct we need to see at least four unfolding peaks of SNase. In the results we have that about 70% of the force extension curves that contain from 4 to up to 6 characteristic unfolding peaks of SNase and that end before reaching the I27 unfolding force. The remaining 30% show I27 peaks and, only 5% of the total unfolding force curves present a force greater than 300 pN, but never higher than 500 pN. Again, as comparison we collect the detachment forces of the $I27 - (Snase - I27)_3$.

It is possible to appreciate in Fig.7.3 that in the construct with SM (blue line) there is a prominence of ruptures at 100 pN that cannot be observed with the $I27_6SM$ construct.

7.1.1   Discussion

How can we interpret the results from the $(I27)_6$ and $(I27 - SNase)_3$ handles? Actually the structural studies of streptavidin tetramers reveal a particular conformation of this object. For the clarity of our exposure we will name the monomers: A,B,C,D.

The studies [84] [83] showed that the monomers associate as dimers with a strong interaction (AB, CD) with a large contact area (approximatively 16

Figure 7.4: In the sketch we can see how four molecules of streptavidin assemble. The strongest interaction is between the monomers AB and CD, then we have weaker interaction (cross and lateral) that appear between the two dimers.

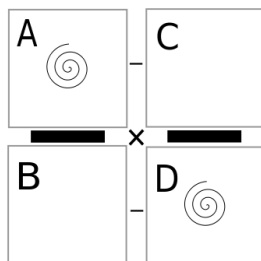nm^2) that involves seventeen hydrogen bonds and other secondary interactions.

The difference in the PDF between the $I27_6SM$ and $(I27 - SNase)_3SM$ can be now be explained. The results in fact suggest that when we are using the I27 handles, we are mainly observing the stronger interaction (between AB or CD) composed by seventeen hydrogen bonds, with an unfolding force between 300 and 500 pN. When we use the $(I27 - SNase)_3SM$, all possibles interaction in the tetramer are explored and, in this case, we also observe the dimer-dimer interaction that consists of 4 hydrogen bonds. In this case we have a much lower rupture force that is consistent with a force of 100 pN. Moreover the PDF suggests that the rupture of the dimers occur more frequently than the strong interaction ruptures (AB, CD).

7.2 Strep-tag Strep-Tactin

Another system that has been studied using the multimodular protein as handles is the Strep-Tactin (an engineered streptavidin) vs Strep-tag II (a particular peptide, the detail of this molecule can be found [85]). This construct have some similarity with the $(I27 - SNase)_3SM$, in this case Strep-Tactin forms tetramers of the same kind of the streptavidin. We can then refer to the Fig. 7.4 to understand the dimer-dimer interactions.

In this case the Strep-tag was fused with $(I27 - SNase)_3$ construct at the C terminus. We then make react this last molecule with the tetramer and, at the adequate concentration, we obtain a structure composed by four handles that bind the four different molecules of the tetramer.

The idea in this case is to have a self assembled tetramer of Strep-Tactin where four different $(I27 - SNase)_3 - Strep - tag$ can bind. Similarly to the case $(I27 - SNase)_3SM$ tetramer, also in this case, to be sure that we are pulling the Strep-tag Strep-Tactin bond we need to see, at least, four SNase unfolding peaks. Experimental results showed that when stretching the Strep-tag and the Strep-Tactin complex, no I27 unfolding force peaks were recorded, this is a strong evidence of the fact that the compplex broke before the I27 rupture
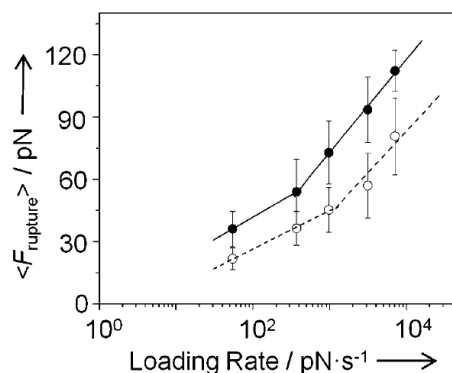
Figure 7.5: In the figure we can see the results for the different loading rate. Each couple of points, at the same loading rate, come from the deconvolution of a single distribution of probability.

force. However, when we tested the handles (without the complex) the peaks of I27 were recorded. Also in this case the PDF of $(I27 - SNase)_3$ and of $(I27 - SNase)_3 - Strep\text{-}tag - Strep\text{-}Tactin - (I27 - SNase)_3$ where compared. The results showed that in presence of the complex we have a distribution of rupture forces that is much wider compared to the distribution of detachment forces of the (just) handles. This rupture forces distribution can be deconvolved in two other distributions, one centered at 48 pN and another at 78 pN (at 1000 nm/s speed). To characterize these interactions we have performed the experiment at different loading rate (from 50 pN/s to 7000 pN/s). These two kinds of interactions can be interpreted by two different linear regimes, however they show always a difference that become more pronounced at higher load (as it is possible to see in Fig. 7.5).

The two interactions can be explained supposing different pulling geometries.

In the case where the interaction is stronger, we can say that we are measuring the interaction of a single Strep-Tactin dimer (similarly to the streptavidin tetramers we are pulling the dimer AB or CD). On the other case, when the interaction is weaker, we are pulling the complex from two different dimers (in this case, using again the sketch of the streptavidin, we are pulling AB and CD and we are feeling the force between these two dimers).

7.3 Angiogenin RNase inhibitor

The handle has been used also to study the Angiogenin vs RNase Inhibitor interaction. In this case $(I27 - SNase)_3$ was fused at the C terminus with RNase and, $I27 - (I27 - SNase)_3$ was fused at the C terminus with Angiogenin. The KDE of the events obtained at the speed of 500 nm/s ranges from few pN to 300 pN. Two humps can be observed in the distribution, one at 78 pN and another at 158 pN. Also in this case the KDE of the Angiogenin vs RNase experiment was

compared with the KDE of the detachment of the handles. The detachment
events show a wider distribution of forces, ranging from few pN to 800 pN.
The two humps of the Angiogenin vs RNase interaction can be explained by
a two-step binding mechanism[86]. It has been reported that RNase inhibitor
and angiogenin may undergo conformational selection by first forming a loosely
bound complex with K_D of 0.5 μM and then forming a tight complex with a
dissociation constant K_D of 0.7 fM. Therefore, the two distributions that we
observed possibly reflect the rupture forces of the weak and tight complexes
with different binding affinities of the RNase inhibitor/angiogenin complexes.

Chapter 8

Cantilever properties

8.1 Autocorrelation and convolution

In this chapter we are going to show the result of two mathematical analysis done on several cantilevers by two mathematical methods: Fast Fourier Transform (FFT) and autocorrelation. These two methods has been used, concurrently with a theoretical model that describe the cantilever as a overdamped oscillator, to obtain the cantilever mass and the viscous drag.

This have been done to confirm the previous results on this topic regarding the properties of a system under external modulation. Several experiments have also been performed to study this topic more in depth, however the results are still extremely challenging to interpret and, for such reason, they will be not shown here.

Before introducing the autocorrelation analysis results it is worth to introduce some aspect of this mathematical operation.

Most importantly, autocorrelation is a mathematical operation based on the another operation, namely convolution. If the convolution between two signals is defined as eq. 8.1 (also called cross-correlation):

$$C(t) = \int_0^t \phi(s)\psi(t-s)ds \qquad (8.1)$$

for $0 \leq t < \infty$.

For discrete signal we have eq. 8.2.

$$C(\tau) = \sum_{n=n_0}^{n_0+m} \phi(n)\psi(n-\tau) \qquad (8.2)$$

Convolution has been used by the author as an algorithm in [37], where an "L"-shaped vector was convolved with the signal obtained from the AFM to localize the peaks of the force spectroscopy signal of the unfolding of multimodular proteins. In such case, when an unfolding and the successive relaxation
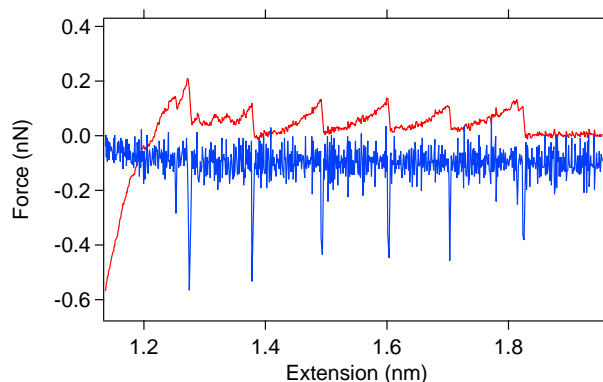
Figure 8.1: In red we have a signal of a multimodular protein (in particular it is a fragment of $(I27)_8$). The blue signal is instead the result of the convolution of the original signal with the "L" shape vector.

occur, the shape of the relaxation signal looks like an "L" (see Fig. 8.1). Convolving this region with an "L" vector yields a sharp spike in the convolved signal. The convolved signal is relatively flat but it show spikes in the regions where we have unfolding signals. This characteristic allows us to analyze the signal just by looking at the evident spikes corresponding to the occurrence of the "L" shape, allowing their recognition to be much easier than by directly looking at the original instrumental signal.

Finally, the convolved signal can be analyzed to localize the peaks with statistical methods.

8.2   Autocorrelation

We can imagine to convolve a (smaller) section of a signal with a larger section of itself, this operation is called autocorrelation. The procedure consist in taking a signal of length "N" and calculate the convolution for "l" value (called also lag).

The mathematical definition of discrete and unitary convolution is given in equation 8.3.

$$R(l) = \frac{1}{(N-l)\sigma^2} \sum_{t=1}^{N-l} (X_t - \mu)(X_{t+l} - \mu) \tag{8.3}$$

The autocorrelation function is a mathematical tool similar to the Fourier Transform (FT), however FT investigates the signal in the frequency domain, instead the autocorrelation is a representation of the time domain. These two tools are bounded together by the Wiener-Khinchin theorem that states that: "the power spectral density of a wide sense stationary random process is the Fourier transform of the corresponding autocorrelation function."

8.2.1 Spectrum of a cantilever

A cantilever in liquid can be modeled as a 1-d underdamped oscillator satisfing the following Langevin equation:

$$m\ddot{x}(t) = -Kx(t) - \gamma\dot{x}(t) - \xi(t) \tag{8.4}$$

Where: "K" is the spring constant, "γ" is the damping coefficient and the therm ξ is a brownian white noise that satisfing the following conditions:

$$\langle\xi(t)\rangle = 0$$
$$\langle\xi(t)\xi(t')\rangle = 2m\gamma K_B T\delta(t - t') \tag{8.5}$$

If we use the Fourier Transform we have:

$$(-m\omega^2 + i\omega\gamma + k)x(\omega) = F(\omega) \tag{8.6}$$

Where $F(\omega)$ is the Fourier Transform of the noise. We can rearrange eq.8.6 to express the power spectrum:

$$|x(\omega)|^2 = \frac{|F(\omega)|^2/m^2}{(\omega_0^2 - \omega^2)^2 + \beta^2\omega^2} \tag{8.7}$$

Where we have introduced the following notation: $k/m = \omega_0^2$ and $\gamma/m = \beta$. Using a corrolary of the Parseval theorem we can write:

$$PSD(x,\omega) = \lim_{t_T\to\infty}\frac{1}{t_T}\frac{2|F(\omega)|^2/m^2}{(\omega_0^2 - \omega^2)^2 + \beta^2\omega^2} \tag{8.8}$$

The thermal noise in this kind of oscillator is assumed to be white, so we can write one part as $D_0 = \lim_{t_T\to\infty}\frac{1}{t_T}|F(\omega)|^2/m^2$, and plugging it into eq. 8.8 we can write:

$$PSD(x,\omega) = \frac{D_0}{(\omega_0^2 - \omega^2)^2 + \beta^2\omega^2} \tag{8.9}$$

We can express D_0 but to do this we need to go further, first we integrate the eq.8.9.

$$\int_0^\infty PSD(x,\omega)d\omega = \frac{D_0}{2m^2}\int_0^\infty\frac{1}{(\omega_0^2 - \omega^2)^2 + \beta^2\omega^2}d\omega = \frac{D_0\pi}{2m^2\beta\omega_0^2} \tag{8.10}$$

Using the definition of ω_0 we can write $\frac{D_0\pi}{2m\beta k}$.

As a corollary of the Parsevel theorem [13] we have $\langle x(t)^2\rangle = \frac{D_0\pi}{2m^2\beta\omega_0^2}$. By using the equipartition theorem we have $\frac{1}{2}\langle x^2\rangle = \frac{1}{2}k_B T$. Using the previous definitions $\frac{D_0\pi}{2m^2\beta\omega_0^2} = k_B T$ and thus $D_0 = \frac{2\beta m k_B T}{\pi}$.

Finally, putting D_0 in eq. 8.9 we have:

$$PSD(x,\omega) = \frac{2\beta k_B T}{\pi m((\omega_0^2 - \omega^2)^2 + \beta^2\omega^2)} \tag{8.11}$$

Eq. 8.11 is a Lorentzian function in ω and it has been used as fitting function to obtain the physical parameters of several cantilevers from the power spectrum.

8.2.2   Autocorrelation of cantilever oscillation: theory

Multipying eq. 8.4 by $x(0)$ and taking the time average we have:

$$m\ddot{C}(t) = -KC(t) - \gamma\dot{C}(t) \tag{8.12}$$

where $C(t) = \langle x(t)x(0) \rangle$.

To solve this equation we use the following ansatz:

$$C(t) = A\cos(\omega t + \phi)exp(-\frac{t}{\tau}) \tag{8.13}$$

Inserting eq.8.13 in eq.8.12 we obtain:

$$\begin{cases} C(t) = A\cos(\omega t + \phi)exp(-\frac{t}{\tau}) \\ \dot{C}(t) = \frac{1}{\tau}A\cos(\omega t + \phi)exp(-\frac{t}{\tau}) - \omega A\sin(\omega t\phi)exp(-\frac{t}{\tau}) \\ \ddot{C}(t) = \frac{1}{\tau^2}A\cos(\omega t + \phi)exp(-\frac{t}{\tau}) - \omega^2 A\cos(\omega t + \phi)exp(-\frac{t}{\tau}) \\ +2\frac{\omega}{\tau}A\sin(\omega t + \phi)exp(-\frac{t}{\tau}) \end{cases} \tag{8.14}$$

Using the expression for $C(t)$ it is possible to simplify the set of eq.8.14.

$$\begin{cases} \dot{C}(t) = \frac{1}{\tau}C(t) - \omega A\sin(\omega t + \phi)exp(-\frac{t}{\tau}) \\ \ddot{C}(t) = \frac{1}{\tau^2}C(t) - \omega^2 C(t) + \frac{2\omega}{\tau}A\sin(\omega t + \phi)exp(-\frac{t}{\tau}) \end{cases} \tag{8.15}$$

Equation 8.12 become:

$$\left(\frac{1}{\tau^2} - \omega^2\right)C(t) + \frac{2\omega}{\tau}A\sin(\omega t + \phi)exp(-\frac{t}{\tau})+$$
$$-\frac{\gamma}{\tau m}C(t) - \omega A\sin(\omega t + \phi)exp(-\frac{t}{\tau}) + \frac{K}{m}C(t) = 0 \tag{8.16}$$

Grouping the terms with $C(t)$ and $A\sin(\omega t + \phi)exp(-\frac{t}{\tau})$ we obtain two terms which must be zero:

$$\begin{cases} \left(\frac{1}{\tau^2} - \omega^2 - \frac{\gamma}{\tau m} + \frac{K}{m}\right)C(t) = 0 \\ \left(\frac{2}{\tau} - \frac{\gamma}{m}\right)A\sin(\omega t + \phi) = 0 \end{cases} \tag{8.17}$$

leading to the following equations:

$$\begin{cases} \frac{1}{\tau^2} - \omega^2 - \frac{\gamma}{\tau m} + \frac{K}{m} = 0 \\ \frac{2}{\tau} - \frac{\gamma}{m} = 0 \end{cases} \tag{8.18}$$

Solving the system we get:

$$\begin{cases} \tau = \frac{2m}{\gamma} \\ \omega = \sqrt{\frac{K}{m} - \frac{\gamma^2}{4m^2}} \end{cases} \tag{8.19}$$

In order to fix the phase ϕ we resort to the virial theorem. So we have:

$$\langle \dot{p}(0)x(0) \rangle = -K_B T \tag{8.20}$$

Substituting $\dot{p}(0)$ with $m\ddot{x}(0)$ and resorting we have:

$$\langle \ddot{x}(0)x(0) \rangle = -\frac{K_B T}{m} \tag{8.21}$$

Since $C(t) = \langle \ddot{x}(0)x(0) \rangle$, we also have $\ddot{C}(t) = \langle \ddot{x}(t)x(0) \rangle$ and therefore (using eq.8.13):

$$\ddot{C}(0) = \langle x(0)x(0) \rangle = A\cos(\phi) \left[\frac{1}{\tau^2} - \omega^2 + \frac{2\omega}{\tau}tan(\phi) \right] \tag{8.22}$$

Now, by equipartition we have:

$$\langle x(0)x(0) \rangle = \langle x^2 \rangle = A\cos(\phi) = \frac{K_B T}{K} \tag{8.23}$$

and therefore:

$$\ddot{C}(0) = \frac{K_B T}{K} \left[\frac{1}{\tau^2} - \omega^2 + \frac{2\omega}{\tau}tan(\phi) \right] \tag{8.24}$$

Using eq. 8.21 and eq.8.24 we have:

$$tan(\phi) = \frac{1}{2} \left[\omega\tau - \frac{1}{\omega\tau} - \frac{K\tau}{m\omega} \right] = -\frac{\gamma}{\sqrt{4Km - \omega^2}} \tag{8.25}$$

Now we can take these results and put them in the ansatz:

$$C(t) = A\cos\left(\sqrt{\frac{K}{m} - \frac{\gamma^2}{4m^2}}t + atan(-\frac{\gamma}{\sqrt{4Km - \omega^2}}) \right) exp\left(-\frac{\gamma t}{2m} \right) \tag{8.26}$$

Eq. 8.26 have been used to fit the autocorrelation spectrum obtained from different cantilever.

8.2.3   Autocorrelation of cantilever oscillation: results

Triangular Veeco cantilever (DNP) with a nominal spring constant of 0.06 N/m and rectangular MikroMasch cantilever (CSC38) have been used. The experiments have been performed in a Veeco AFM (Picoforce-Nanoscope III) with a fluid cell chamber using the "o-ring". In the experiments that we have performed we have recorded the deflection signal coming from the cantilever placed in water at several μm of distance from a surface of mica. Signal Access Module

III (Veeco National Instruments) has been used to acquire the deflection signal, the signal have then been recorded in a home-made software and then analyzed with "Xmgrace".

The typical deflection signal versus time can be seen in Fig. 8.2.
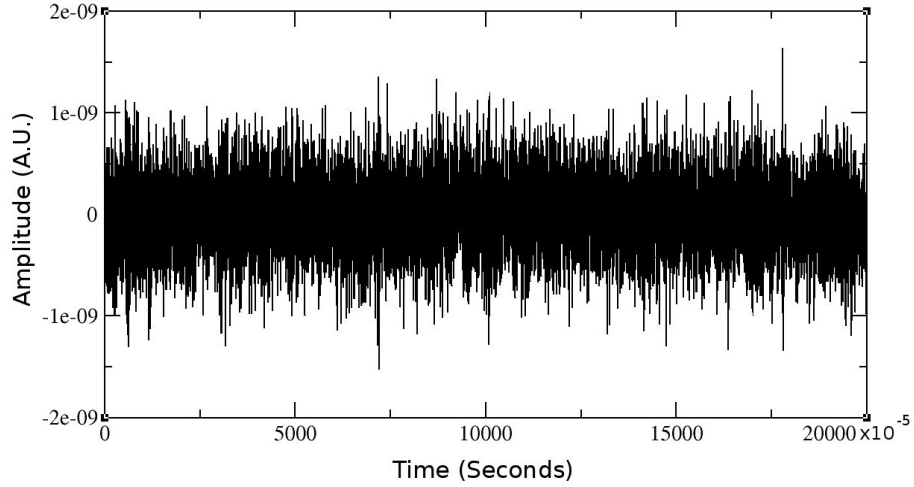
Figure 8.2: Deflection signal of a triangular cantilever in water. The sampling rate was 100 kHz, this mean that the timescale, in this case, is $(1/100)kHz^{-1}$. In this case the duration of the signal was 0.2 seconds (20000 points).

In Fig. 8.3 we can see the autocorrelation function for the deflection signal of the Fig. 8.2 that lasts 0.2 seconds.

In the first attempts to get the autocorrelation plots we have noticed that a periodic correlation appears at high lag value, like in Fig.8.3 and Fig. 8.4 (this latter autocorrelation plot does not came from the same cantilever as before, but the periodic correlations are present in all the autocorrelation plots where the signal duration is in the order of few seconds). We found that the sources of the correlations at high lag are due to the small recording time of the signal and to investigate it better we have acquired signal of different duration from the same cantilever. Using these latter signals we have then generated the autocorrelation plots.

We found that to eliminate the correlation at high lag we have to use a signal that is, at least, 60 seconds long.

In such case the average value of the correlation increases (from "0" to values near 0.9) and the amplitude of the oscillations decrease. This is due to the fact that to correctly average the noise we need a longer time, however in such case the low frequency noise appear. Nevertheless this low frequency noise does not disturb the fitting, in fact the results of the fitting using the model 8.26 (where we have put a constant D_0) did not change.

We can try to obtain the values of γ and the effective mass from the power
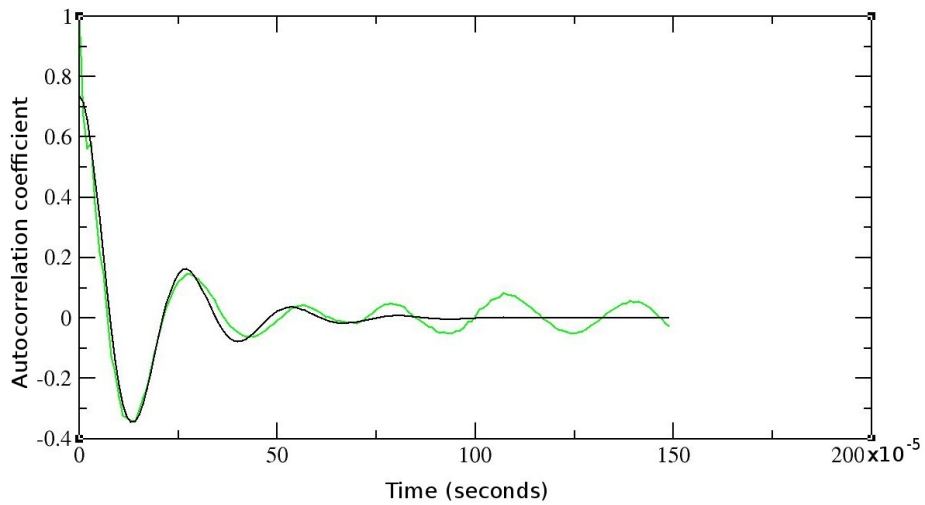
Figure 8.3: The green line represent the autocorrelation function for the signal in Fig.8.2. The black line is the fit using eq.8.26, the results of this fit gave a mass of 66 pg and a γ of $8.9*10^{-7}$ Ns/m.
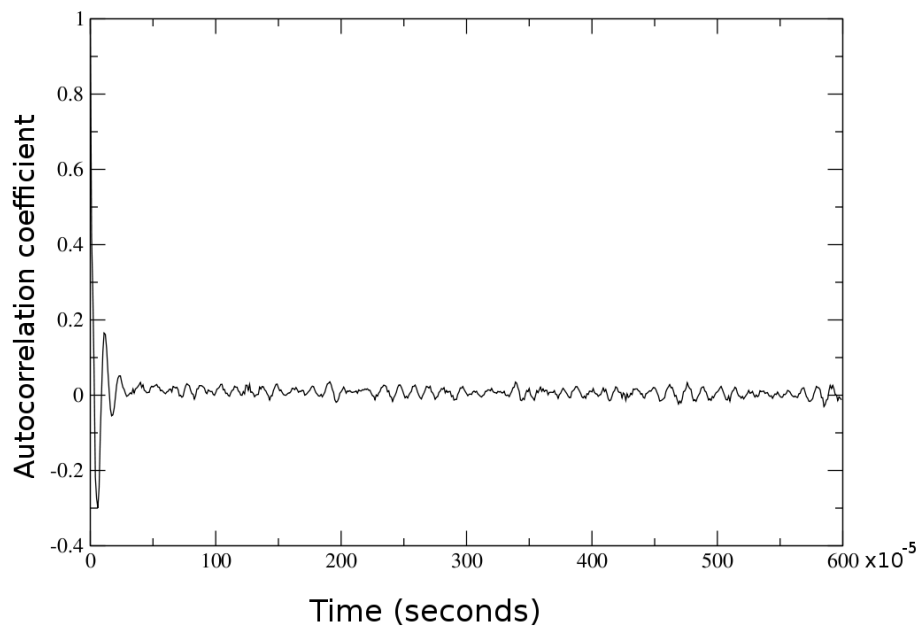
Figure 8.4: Graph for the autocorrelation function on 600 points. Here the signal length was 200 ms, the sampling rate was 100 kHz.
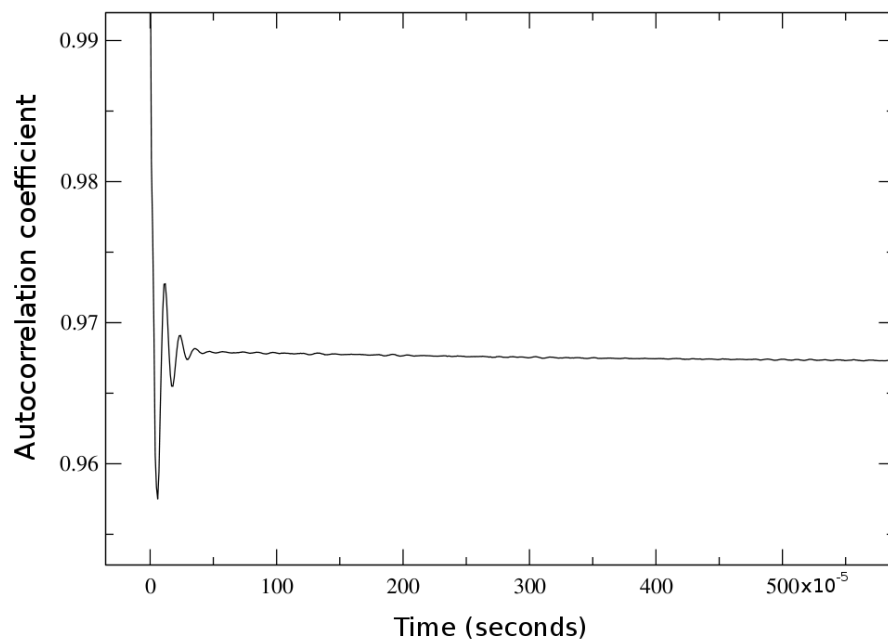
Figure 8.5: The same signal from Fig.8.4 was used but here the time length was 68 s, the sampling rate was 100 kHz. It is possible to see that the value of baseline of the autocorrelation function increase compared to the baseline in Fig.8.4 and that the oscillation at high lag disappear.

spectrum of the cantilever. The power spectrum can be approximated by eq. 8.27.

$$< p^2(\omega) >= \frac{2\gamma^* k_B T}{(K_{cantilever} - m^*\omega^2)^2 + (\gamma^*\omega)^2} \qquad (8.27)$$

However this model is not precise enough to describe the entire spectrum and it has to be restricted to the resonance peak. Moreover a three parameter fit can be used to better fit the spectrum, the equation for this is eq.8.28.
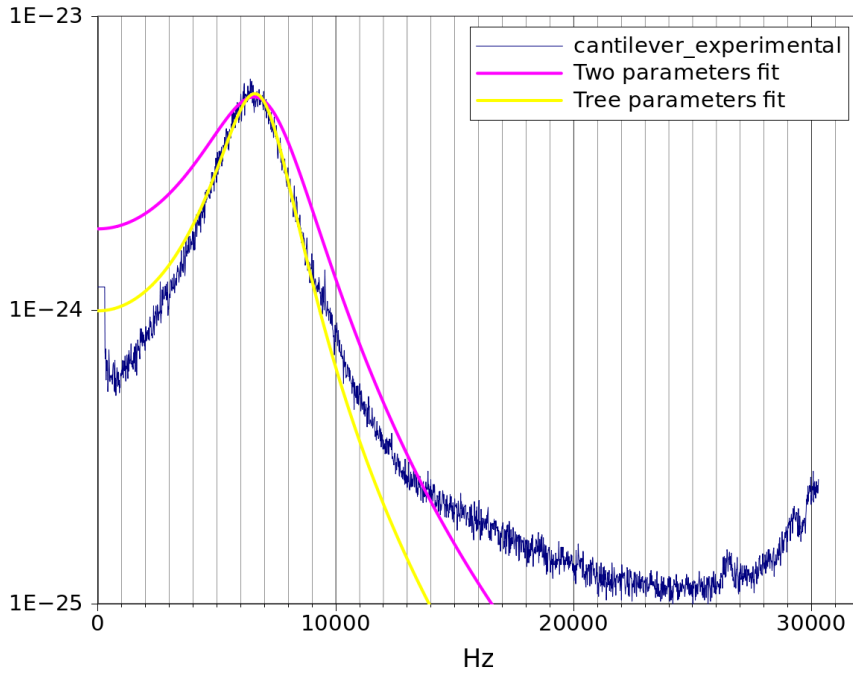
Figure 8.6: Fit of the power spectrum of a cantilever in water with eq. 8.27 (violet line) and 8.28 (yellow line). The parameters are, respectively mass 28 pg and γ $8.2*10^{-7}$ Ns/m for the two parameter fit and 31 pg and γ $6.0*10^{-7}$ Ns/m for the three parameters fit.

$$< p^2(\omega) >= \frac{A}{(K_{cantilever} - m^*\omega^2)^2 + (\gamma^*\omega)^2} \qquad (8.28)$$

Here we can see the results of the various fits done with different cantilevers and different methods. It is possible to see that the values obtained are in good agreement, this demonstrates the reliability and the consistency of the methods used to obtain such parameters.

Summary			
Cantilever	3 parameter fit	2 parameter fit	Autocorrelation fit
20100309 (1)	A0=1.48e-26 m=6.92E-11 $\gamma = 1.03E-6$	ind m=7.89E-11 $\gamma = 7.07E-7$	A0=0.75 m=8.38E-11 $\gamma = 9.45E-7$
20100323 (2)	A0=1.79e-26 m=7.50E-11 $\gamma = 1.29E-6$	ind m=8.72E-11 $\gamma = 9.26E-7$	A0=0.73 m=8.93E-11 $\gamma = 1.04E-6$
20100324 (3)	A0=1.57e-26 m=6.85E-11 $\gamma = 1.2E-6$	ind m=7.88E-11 $\gamma = 8.93E-7$	A0=0.83 m=7.16E-11 $\gamma = 1.61E-6$
20100412_1 (4)	A0=1.23e-26 m=4.45E-11 $\gamma = 8.90E-7$	ind m=5.47E-11 $\gamma = 6.43E-7$	A0=0.75 m=5.44E-11 $\gamma = 9.52E-7$
20100412_2 (5)	A0=1.48e-26 m=5.51E-11 $\gamma = 1.07E-6$	ind m=6.70E-11 $\gamma = 7.73E-7$	A0=0.79 m=7.03-11 $\gamma = 1.15E-6$
20100412_3 (6)	A0=1.21e-26 m=4.52E-11 $\gamma = 8.50E-7$	ind m=5.57E-11 $\gamma = 5.88E-7$	A0=0.79 m=4.72E-11 $\gamma = 1.01E-6$
20100412_4 (7)	A0=1.76e-26 m=6.00E-11 $\gamma = 1.21E-6$	ind m=7.47E-11 $\gamma = 8.48E-7$	A0=0.83 m=6.67E-11 $\gamma = 1.41-6$
20100412_5 (8)	A0=1.84e-26 m=6.99E-11 $\gamma = 1.28E-6$	ind m=8.30E-11 $\gamma = 9.00E-7$	A0=0.82 m=8.11E-11 $\gamma = 1.63E-6$
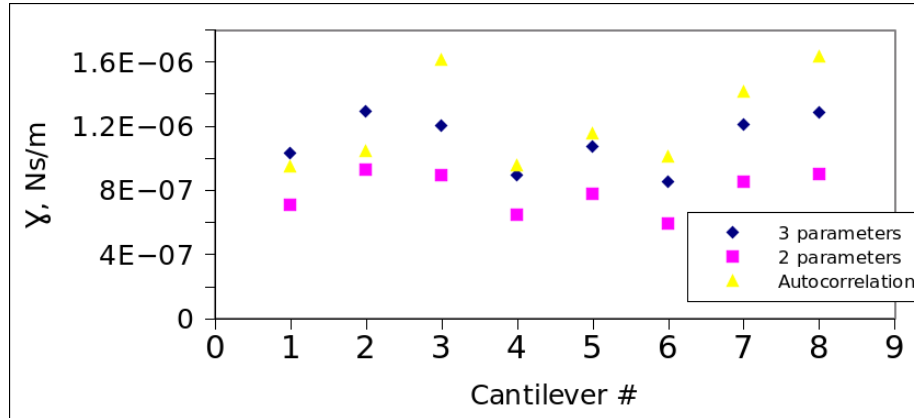
Figure 8.7: The values for the γ estimated with the three different methods is represent in figure.
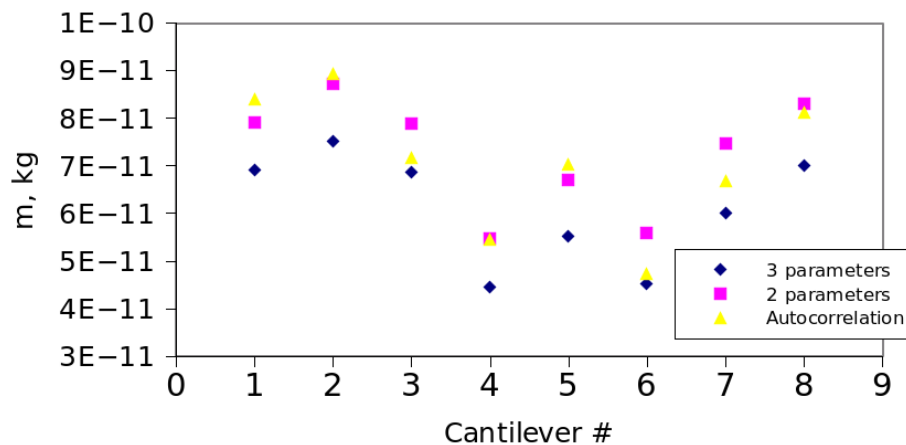
Figure 8.8: The values for the mass estimated with the three different methods is represent in figure.

8.3 Conclusions

According to the Fig.8.7 the fits performed with the power spectrum with three parameters and the autocorrelation are more close compared to the 2 variable fit.

We can also notice that the mass estimated with the autocorrelation is always larger that the mass estimated with the three variable fit, moreover this parameter is really close to the mass estimated with the two variable fit.

Speaking about the estimation of γ (Fig.8.7) we can see that, in the set of measurements, the autocorrelation function gives most often larger values, however it is not always true. In the two graphs we can see that the values are compatible with a range of $2*10^{-7}Ns/m$ for γ and 10^{-11} kg for the mass. This kind of results are fundamental for the interpretation of the experiments where a small modulation at high frequency is used when a polymer is pulled (see in particular Favre et Al. [69] and Chtcheglova et Al. [87]).

Chapter 9

Dithering technique

The elastic properties of the biomolecules have been investigated deeply with the use of standard AFM methods. The main protocol has been the velocity clamp on one hand because of technological simplicity but most importantly because, in contrast with the force clamp, it allows to measure directly the force response of a biomolecule in function of its extension.

It is exactly the force extension profile that allows us to calculate the stiffness as the derivative of the force as a function of the elongation: this is impossible if we use the force clamp (or force ramp) techniques.

The velocity clamp allows us, as saw in the chapter 3, after some analysis, to measure different properties of a polymer, like the stiffness of a polymer and also of the folded modules in a multimodular protein.

However the simple velocity clamp consists of the measurement of the static force extension profile of a polymer. For a better understanding of the process that are involved in the polymer mechanics it is important to have an estimation of the dynamic parameters like the dynamic stiffness and damping behavior.

Such parameters could be mathematically obtained from the static force extension curve measuring the slope of a curve in a determine point, however the methodology that we are going to introduce is able to direct obtain this value.

These measurements can be performed applying a modulation signal to the cantilever base. The signal is then processed by a lock-in to get the changes of the amplitude and phase of the response.

This kind of measurement are still at the beginning even important works like [15] have been already published. Here we would like to continue the work started with Chtcheglova [87] and Favre [69], however in our case we have decide to modulate the cantilever base at low frequency.
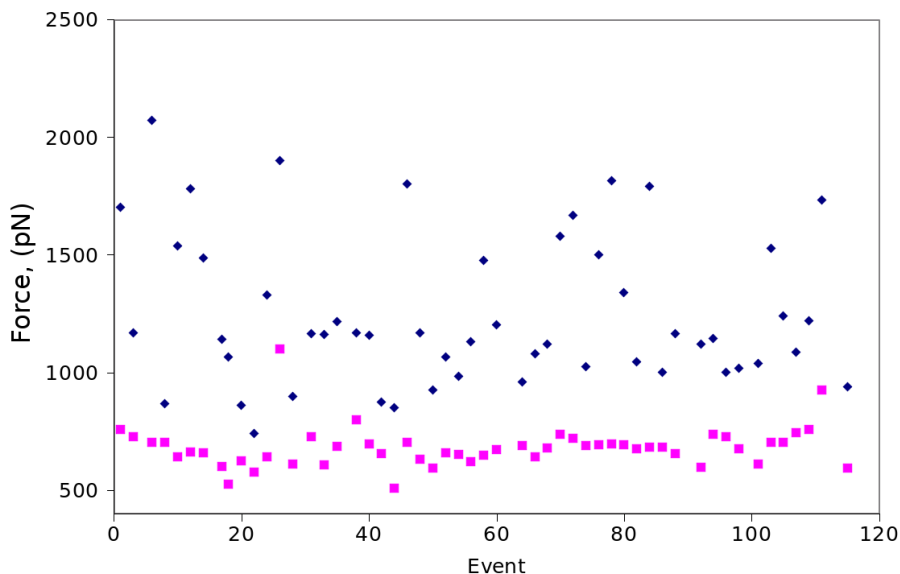
Figure 9.1: Several curves of dextran have been analyzed. We we have measured the force of the transition region taking a point in the middle of the transition region, before the detachment (violet square). The blu diamonds instead represent the detachment force.

9.1   Preliminary results

Before introduce the modulation technique we want to illustrate some preliminary results that we have obtained. In fact, at that time, we needed a molecule that was easy to store, cheap and easy to obtain. Thanks to the suggestion of Prof. Marszalek we decided to use the dextran. The dextran can be stored as a powdered and, at the usual purity that we need it is not expensive. We did not have any experience with this molecule and we had to develop our technique of deposition. We then tried to use a glass coverslip cleaned with the plasma to prepare the samples. The functionalization of the coverlisp, using the dextran, is also discuss in "Introduction".

During the preliminary tests we were able to record several force curves. To check the conditions we have also measured the force at which the transition region occur, such result can be seen in Fig. 10.1. These firsts results were in agreement with the values obtained in literature and allowed us to continue the work to test the modulation technique.

9.2   Lock-in modulation and signals

To start we have to introduce the Lock-in and the signal processing that this instrument is capable to do. A lock-in measure require an instrument with a

precise internal reference. In general such reference is a periodic function at a fixed frequency (and in our case it is). In the lock-in, in practice, we can measure the response of a setup excited at the reference frequency. Let assume that we have a reference signal with a frequency ω_r and let denote the amplitude of such signal V_l. We can use this signal to excite the base of the cantilever, the response can be think as a waveform. The response signal is $V_{sig}sin(\omega_r t + \theta_{sig})$ where V_{sig} is the value of the amplitude. The lock-in generate the following reference $V_L sin(\omega_L t + \theta_{ref})$.

The lock-in amplifies the signal and then multiply it by the reference. The output signal V_{out} is a product of two sine waves:

$$V_{out} = (1/2)V_{sig}V_L cos([\omega_r - \omega_L]t + \theta_{sig} - \theta_{ref}) -$$
$$(1/2)V_{sig}V_L cos([\omega_r + \omega_L]t + \theta_{sig} + \theta_{ref}) \tag{9.1}$$

The V_{out} is a signal composed by two waves, one at the frequency $(\omega_r - \omega_L)$ and the other at the frequency $(\omega_r + \omega_L)$. If we pass V_{out} in a low pass filter the time dependent part of the signal disappear and, if we have $\omega_L = \omega_r$ we will have: $V_{out} = (1/2)V_{sig}V_L cos(\theta_{sig} - \theta_{ref})$.

This signal is now independent from the time and proportional to the signal amplitude.

Now let assume that the input is composed by signal and noise. The PSD and the low pass filter will detect the signals that have a frequency very close the the reference frequency. The other signals, with a frequency far from the reference, are attenuate by the low pass filter. Noise at a frequency very close to the reference frequency will result in a low frequency signal. Their attenuation depends upon the low pass filter bandwidth and roll-off. A narrower bandwidth will remove noise sources very close to the reference frequency, a wider bandwidth allows these signals to pass.

9.2.1 Amplitude and phase

The V_{out} signal is proportional to $\theta = (\theta_{sig} - \theta_{ref})$. θ is the phase difference between the signal and the lock-in internal reference. By controlling θ_{ref} we can make θ equal to zero and in that case we can directly measure V_{sig}. On the other hand, if θ is 90°there will be not output at all.

This phase dependency can be eliminated multiplying the V_{out} by another wave shifted by 90°and the low pass filtered output will be:
$V_{psd2} = (1/2)V_{sig}V_L sin(\theta_{sig} - \theta_{ref})$, and $V_{psd2} \sim V_{sig}sin\theta$. These two outputs, that we are going to call "X" and "Y", are: $X = V_{sig}cos\theta$ and $Y = V_{sig}sin\theta$.

The magnitude (or amplitude) "R" is $R = (X^2 + Y^2)^{1/2} = V_{sig}$. R then measure the signal amplitude and does not depend on the phase between the signal and the reference.

In addition we can measure also the phase between the signal and the reference and so we have $\theta = tan^{-1}(Y/X)$.

9.3   Theory

For an analysis of the experimental data obtained at low frequencies we tried to use the approach of Burhman et al. [88], the case "displacement modulation via tip" fits to our experimental set up. Using the notation "z", "d" to describe respectively the position of the cantilever base and tip relative to the immovable support (the latter is at the same time the end-to-end distance for the stretched molecule), we write the equation of the motion as

$$m\ddot{d} + 2m\beta_c(\dot{d} - \dot{z}) + k_c(d - z) + 2m\beta_i\dot{d} + g(d) = 0 \qquad (9.2)$$

Here "m" is an effective mass of the cantilever while $\beta_{c,i}$ and $k_{c,i}$ are dissipation and spring constants characterizing respectively cantilever and molecular complex under the study and g(d) is just the force experienced by the stretched molecule. Further, in this equation, $z = z_0 + \delta z e^{i\omega t}$ with $z_0 = vt$ is the initial condition which reflects the imposed slow constant velocity motion of the cantilever base (force loading) and its dithering at frequency ω with an amplitude δz while the solution is searched in the form $d = d_0(t) + \delta d e^{i\omega t}$.

During the procedure of restoration of the static signal we have used equation 9.3 that provided a good approximation of the static signal.

$$S_{st}(t) = \frac{bv}{a} \int_0^t (S_{dyn}(\tau) + c)d\tau \qquad (9.3)$$

9.4   Experimental conditions

Before introducing the results we need to speak about the correct conditions to perform the experiment. The dithering technique require to integrate one or more periods of the exciting signal to correctly obtain the local amplitude and phase. We are substantially play with the following parameters:

- resolution in therm of "integrated periods for nm"

- frequency used,

- pulling speed.

It is easy to understand that an really high frequency require a short integration time in rapport of the time that we need to do a ramp, but how all these parameters are related and how do we fix them before perform an experiment?

The following procedure should be used to obtain the right values of frequency and speed once we have decided the resolution desired.

Let fix the resolution as "a" period/nm. We want also that the frequency is not bigger that "k" Hz. It is straightforward, then, that the maximum speed that we could apply to the experiment is given by: $v = k/a$. That simple equation and this simple reasoning can be reversely used to find the other conditions starting from the speed.
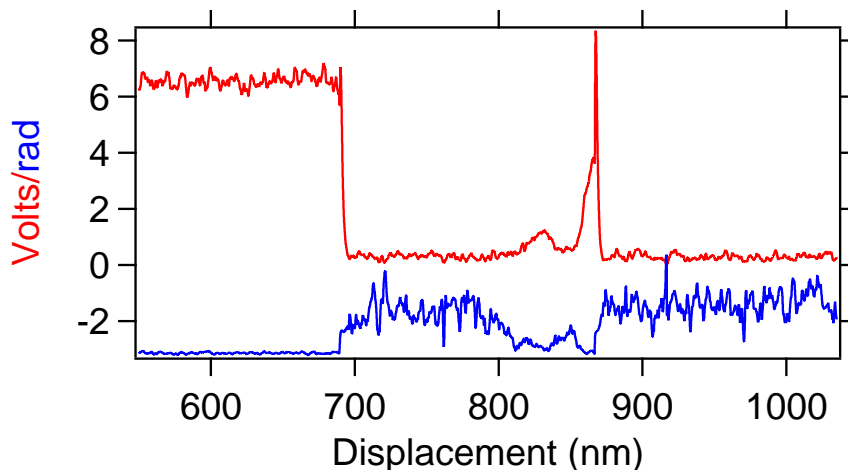
Figure 9.2: In the figure we can observe the amplitude channel (red line) and the phase (blue line). The amplitude correctly reproduce the stiffness of the polymer during the pulling process.

9.4.1 Results

If we respect the conditions for the correct sampling rate we can obtain signals that describe well the force curves. It is possible to see in Fig.9.2 how, with a speed of 25 nm/s, a modulation of 80 Hz of frequency and (considering that the cantilever spring constant was 0.1 N/m) a modulation amplitude of 0.2 nm, we can obtain the signal of amplitude and phase. In this case we proceed to the integration of the amplitude signal with the eq. 9.3, however the integration was done iteratively until a good agreement between the static and the integrated signal was found.

The phase signal also contain information about the molecule, however the noise of this channel is also large. It is possible that, due to the modulation of the base of the chip, the cantilever feel a continuous interaction with the liquid giving rise to the noise recorded in the phase. Other experiments, where the piezo is modulate should be done to confirm this latter hypothesis. Moreover, now that all conditions to perform the experiments are clear it would be interesting to slow down even more the speed to increase the resolution and see if it will be possible to capture more details of the pulling process.
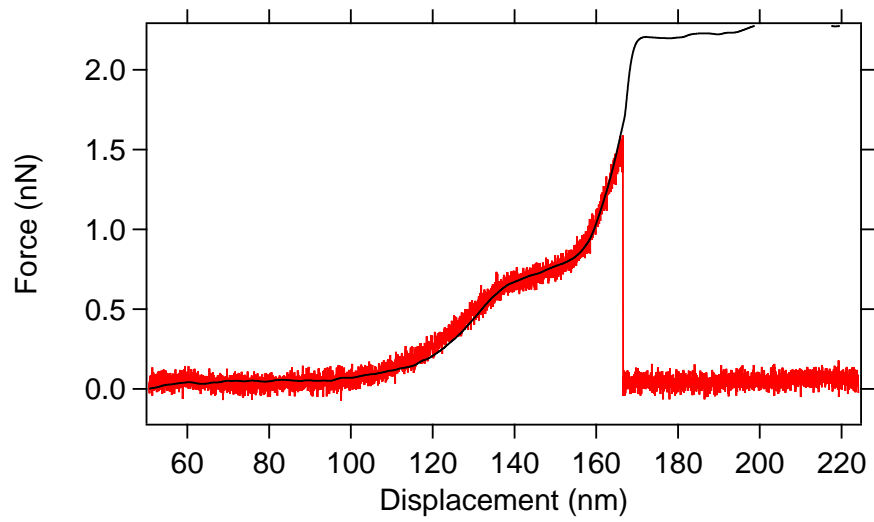
Figure 9.3: In the figure we can see the deflection signal (in red) and the reconstruction with the eq.9.3 (black line).

Chapter 10

Appendix

10.1 Monte Carlo code

Hereafter we can see the Monte Carlo code to generate force unfolding profiles. The code read the parameters of the simulation from a file and then simulate the force extension curve. This code is also highly customizable, the kinetic parameters and the characteristics of the modules can be changed in the "parameters.txt" file. The main difficult is just the cast of the WLC function in a third grade equation, such cast was discuss in chapter 4. When the WLC is in a third grade equation form it can be solved for each "t" to get the current force acting on the system. The current force can then be used to calculate the transition rates to increase, or decrease, the countour length. After have calculate the force and the transition rate we increase the time of the simulation by the timestep, in such way we obtain a new position for the base of the cantilever that mean also a new solution for the WLC.

```cpp
#include <stdlib.h>
#include <iostream>
#include <fstream>
#include <time.h>
#include <cmath>
#include <vector>
#include <sys/time.h>

using namespace std;

// our function is a third grade equation U^3 + bU^2 + cU + d = f(x),
// it represent the WLC(t)
double f(double b, double c, double d, double x){
    return d + x*(c + x*(b + x) );
}
```

```cpp
// first derivative of the function 3U^2 + 2bU + c
double fprime(double b, double c, double x) {
    return c + x*(2*b + 3*x);
}

// it calculate the solutions for the polinomial WLC(t)
bool nr_pol3(double b, double c, double d, double x0, double epsilon,
double& x){
    const int maxIteration = 500;
    x = x0;
    double fx = f(b,c,d,x);

    int numIterations = 0;

    while (fabs(fx) > epsilon) {
        double fpx = fprime(b,c,x);
        ++numIterations;
        if (fabs(fpx)<epsilon)
        return false;
        if (numIterations>maxIteration) { //|| fabs(fpx)<epsilon
//          cout<<"Too many iterations"<<endl;
            return false;
        }
        x -= fx / fpx;
        fx = f(b,c,d,x);
    }
    //cout<< "Number of iterations:"  << numIterations <<endl;
    return true;
}

//if needed we can also calculate the derivative of thw WLC at each
// value of force
double wlcprime(double v,double time,double temp,double kc,double cl,
double pl,double present_force){
//first derivative of the WLC function respect to x
//kbt/p*(  (1/(2*cl*(1- x/cl)^2))  + 1/cl )
double x=v*time-(present_force/kc);
double kbt=1.3806e-23*temp;

return (kbt/pl)*(
    1.0/(  2.0*cl*(1.0-(x/cl))*(1.0-(x/cl))   )  + (1.0/cl)
                                                           );
}

// here we simply calculate the coefficient a,b,c of the equation
//U^3 +aU^2 +bU +c =0 from the parameters
```

```
void coeff(double v,double time,double temp,double kc,double cl,
double pl, double *sol)
{
 double kbol, s, vt, SL, den, svt, b, c, d;
 kbol=1.3806e-23;
 s=kc*pl/(kbol*temp);
 vt=v*time;
 SL=s*cl;
 den=-4*SL-4;
 svt=s*vt;
 b=(9.0+4.0*svt+8.0*SL)/den;
 c=(-6.0-8.0*svt-4.0*SL)/den;
 d=4.0*svt/den;
 sol[0]=b;
 sol[1]=c;
 sol[2]=d;
}

double getFt(double a,double v,double time,double cl,double kc)
{return (v*time-a*cl)*kc;}

// give the probability to unfold/(have a transition) of an
// element given the kinetic parameters
double prob_of_unf_t(double koff,double deltax,double temp,
double timestep,double mod,double F){
 double kbol=1.3806e-23;
 return mod*timestep*koff*exp((deltax/(kbol*temp))*F);
}


// give the probability to fold/(have a transition) of an
// element given the kinetic parameters
double prob_of_fold_t(double koff,double deltax,double temp,
double timestep,double mod,double F){
 double kbol=1.3806e-23;
 return  mod*timestep*koff*exp((deltax/(kbol*temp))* (-F));
}



void sim_unf_wlc(double vel, double kel, double modclength,
double firstclength, double plength,
double deltax, double koff, double kon,
double temp, double timestep, const int nummod,
vector<double> *dev_a, vector<double> *length)
{
```

```c
// here we assigne the inizial contour length given by the length of
// the FOLDED molecule
double rval, clength, tplgth, sim_time, phys_sol_U, abc[3];
clength=firstclength;
sim_time=0.0;
phys_sol_U=1e-12;

// C requires "struct timval" instead of just "timeval"
struct timeval tv;
gettimeofday(&tv, 0);
// use BOTH microsecond precision AND pid as seed
long int n = tv.tv_usec * getpid();

srand(n);
double last_ft;
int numbmod=nummod;
int numfold=0;

while (numbmod>0)
{

//using the tplgth  we can also implement a function of
//(numbmod) for the persistence
tplgth=plength;
//find the coefficients a,b,c of the equation
//u^3 +aU^2+bU+c=0 that are function of the parameters
coeff(vel, sim_time, temp, kel, clength, tplgth, abc);
//find the solutions with Newton solver
nr_pol3(abc[0], abc[1], abc[2], phys_sol_U, 1e-12, phys_sol_U);

// from the solution we found the force applied at the system,
// this passage is due to the formalism that
// we have used to solve (and to express the coefficient)
// of the equation of the third grade
last_ft= getFt(phys_sol_U, vel, sim_time, clength, kel);

//now here we compare a random number with the probability to unfold
rval= rand()/((double)RAND_MAX + 1.0);

if ( rval <=
prob_of_unf_t(koff, deltax, temp, timestep, numbmod, last_ft) )
{
clength+=modclength;
numbmod-=1;
}
```

```
rval= rand()/((double)RAND_MAX + 1.0);
//same for the probability to fold
if (  ((nummod - numbmod) >0) &&
(rval <=
prob_of_fold_t(kon,deltax,temp,timestep,nummod-numbmod,last_ft))  )
{
        clength-=modclength;
        numbmod+=1;
        numfold++;
        cout<<"fold:   "<<numfold<<endl;
}

//we store the datas in the arrays each 100 steps (or more)
if (  int(sim_time/timestep)%100==0)
{
dev_a->push_back(last_ft);
length->push_back(sim_time*vel);
}

//we increase the time
sim_time+=timestep;
}


}


int main(){
clock_t t1, t2;
t1=clock();

vector<double> force;
vector<double> length;


//we read the parameters from a file
ifstream rfile("parameters.txt");

int num_modules;
double vel, kel, modclength, modflength, plength, dx,
koff, kon, temp, timestep, add_clenth;
string buff;

//we read the paramters from a file and we jump the header
for(int i=0; i<12;++i)
```

```
    getline ( rfile , buff , '\n' ) ;
rfile >>dx ;              //1
rfile >>koff ;           //2
rfile >>kon ;            //3
rfile >>vel ;            //4
rfile >>kel ;            //5
rfile >>modclength ;     //6
rfile >>modflength ;     //7
rfile >>plength ;        //8
rfile >>temp ;           //9
rfile >>timestep ;       //10
rfile >>num_modules ;    //11
rfile >>add_clenth ;     //12

cout <<"Delta x nm:   "<<dx<<endl ;
cout <<"Speed nm/s:   "<<vel <<endl ;
cout <<"Module Contour length nm:   "<<modclength<<endl ;
dx∗=1e−9;
vel∗=1e−9;
modclength∗=1e−9;
modflength∗=1e−9;
plength∗=1e−9;
timestep∗=1e−9;
timestep/=vel ;
add_clenth∗=1e−9;

cout <<"timestep : "<<timestep <<endl ;

// initial contour length
double firstclength=add_clenth+num_modules ∗(modflength );

cout <<firstclength <<endl ;

    sim_unf_wlc ( vel , kel , modclength , firstclength , plength ,
    dx , koff , kon ,
    temp , timestep , num_modules ,
    &force , &length ) ;


t2=clock ( ) ;
float diff = (( float )t2 − ( float )t1 )/( double )CLOCKS_PER_SEC;
cout <<"Execution time : "<<diff <<" seconds"<<endl ;


ofstream file ;
file . open ("dataout . txt ") ;
```

```
file <<"Parameters:"<<endl;
file <<"Speed nm/s: "<<1e9*vel<<endl
<<"Spring constant cantilever N/m: "<<kel<<endl<<
"Temperature K: "<<temp<<endl<<"Delta nm: "<<
1e9*dx<<endl<<"Koff s^-1: "<<koff<<endl<<
"Number of modules: "<<num_modules<<endl<<
"Persistence length nm: "<<1e9*plength<<endl<<endl;

file <<"pN    nm"<<endl<<endl<<endl;
for (int i=0; i< force.size();++i){
    file <<force[i]<<" "<<length[i]<<endl;
}
file.close();


}
```

10.2 Back calculation

The Back calculation scheme is based on the bootstrap statistic. In case where the loading rate is low enough and the system behave like harmonic springs it is possible to use the forces for a peak to simulate all the others. The following code read a set of data from a file and the sort them in ascendent order.

As input parameters it require the peak order of the data imput (the total number of peaks is fixed to 6).

To extract a single value of a peak in position "n" using the values of a peak in position "n_0" it is sufficient to follow the following procedure.

First we extract a random number "$r = (N-1)*(1.0-(1-f)^{(}n_0))$, where "N" is the number of data points and "f" is a random number in the domain $[0, 1]$. This value "r" is then rounded to an integer "j" and again it is used to calculate the value "$x = r - j$". "j" is then used to find the correspondent force value in the sorted data and "x" is used to calcolate the linear interpolation between the value "j+1" and the value "j".

We repeat it "n" times and we take the minimum of the results obtained.

The aforementioned procedured generate a single value for a peak in position "n", however to have a statistic (as in the code) we have to repeat it several times and calculate the average.

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <vector>
#include <iostream>
#include <fstream>
#include <time.h>
```

```cpp
#include <algorithm>

using namespace std;

int main(){

  vector <double> f;
  double y; // here we store some temporaney data

  double r, x, min, avg, avg2, sigma;
  int j;
  srand(time(NULL));

//Get input,   it have
ifstream infile("datain.txt");
while(!infile.eof())
{
infile >> r;
f.push_back(r);
}

sort(f.begin(),f.end());   //sort the vector in a growing way

infile.close();

int n0; // peak order of the data
int ndat; // number of data points

cout<<"Peak number of data? (considering 1 near the detachment)"<<endl;
cin>>n0;

ndat=f.size();


int num_trials=10000;
 for(int n=1; n <=6 ; n++)
 { // cycle on the peak position
 avg=0.0;
 avg2=0.0;
        for(int k=1; k <= num_trials; k++)
        { //set a num_trials to calculate average and sigma

        min=10000.0;
        for(int l=0; l < n; l++)
        { // we take the minimum on n extraction
          r = (ndat-1.0)*(1.0-pow( 1.0-rand()/float(RAND_MAX),n0));
```

```
        j = (int) (r);
        x = r−j;
        // the value of the extracted force is calculated using a linear
        // interpolation between the values of the force
        y=f[j]+ x*(f[j+1]−f[j]);
      // we take the minimum of the "n" extraction
      if (y < min) min=y;
      }

      avg+=min;
      avg2+= min*min;
      }
avg=avg/num_trials;
avg2=avg2/num_trials;
sigma = sqrt (avg2 − avg*avg);

 printf("N_peaks: %3d   avg_f: %8.3lf   sigma_f: %8.3lf\n",n,avg,sigma);
 }

}
```

10.3 Energy barrier details

Introducing $\varepsilon = 1+\beta F$, i. e. $F = \frac{\varepsilon-1}{\beta}$, we converts (eq. 5.3) into $\lambda = A\varepsilon exp(\tilde{B}\varepsilon)$ with $\tilde{B} = \frac{x_0}{\beta k_B T}$; of course, $A = k_{off}exp(-\tilde{B})$. Let us first consider the case when β is negative so always $\varepsilon \leq 1$. Then it is instructive to introduce the positive constant $B = \frac{x_0}{|\beta|k_B T}$ and write a dissociation law as $\lambda = A\varepsilon exp(-B\varepsilon)$ which is exactly the particular case a=b=1 of Gargs law $\lambda(\varepsilon) = A\varepsilon^{a+b-1}exp(-B\varepsilon^b)$. His consideration briefly goes on as follows[67]. First we calculate the probability function 5.5 which in terms of ε is equal to

$$p(\varepsilon)d\varepsilon = \lambda(\varepsilon)\frac{1}{\dot{\varepsilon}}exp(-\int_\varepsilon^1 \frac{\lambda(\varepsilon')}{|\varepsilon'|})d\varepsilon \qquad (10.1)$$

For the value of $s = \frac{1}{|\dot{\varepsilon}|}\int_\epsilon^1 \lambda(\varepsilon')d\varepsilon'$ Garg notes that for large B (i.e. for our case when $\beta << \frac{x_0}{k_B T}$) one can extend the limit of integration from 1 to infinity to get $s = \frac{A}{|\dot{\varepsilon}|B^2}exp(-B\varepsilon)(B\varepsilon + 1)$ and thus

$$lns = ln\frac{A}{|\dot{\varepsilon}|B^2} - B\varepsilon + ln(B\varepsilon) + ln(1 + \frac{1}{B\varepsilon}) \qquad (10.2)$$

(The calculation of s is, of course, the main hint of Garg: he noted that, due to $dexp(-s) = exp(-S)\frac{ds}{d\varepsilon}d\varepsilon$, the averages $\langle\varepsilon^2\rangle = \int_0^\infty \varepsilon^2 p(\varepsilon)d\varepsilon$, is nothing else than $\langle\varepsilon^n\rangle = \int_{\varepsilon=0}^{\varepsilon=\infty} \varepsilon^n d(exp(-s)) \cong \int_0^\infty \varepsilon^n(s)exp(-s)ds$ hence all one needs is to express ε as a function of s. Note also the "automatic" normalization appearing

here: $\int_0^\infty exp(-x)dx = 1$. The next step is the use of Taylor expansion for an analysis of 10.2 in the following approximation:

- 1) important values of s are of the order of unity,

- 2) $B\epsilon >> 1$,

- 3) $\frac{A}{|\dot\varepsilon|B^2} >> 1$ (note that when s is around 1, $ln\frac{A}{|\dot\varepsilon|B^2}$ must be positive due to negativity of $-B\varepsilon$).

The first condition is evident simply because otherwise the factor exp(-s) is too small to contribute essentially to the average values. The third condition can be rewritten as $\dot F << k_{off}B^{-2}exp(B)/|\beta|$ hence its validity depends first of all on the value of k_{off}. Remind that we also have an opposite "natural" dynamic force spectroscopy limit discussed above. For large B we have $exp(B)/B >> 1$ hence evidently there is a sufficient room for the validity of this condition. Taking as an example the same B=10, effective barrier thickness x_0 around 0.1 nm and the room temperature case $k_BT \cong 4 \cdot 10^{-21}$ J, we have that it should be $\dot F << 10^{-7} \cdot k_{off}$ ($\dot F$ is measured in N/s and k_{off} in s^{-1}) which can be sometimes problematic for k_{off} of the order of a few s^{-1} and less but definitely holds for k_{off} or the order of 100 s^{-1} and larger. The most stringent criterion is the second one which for our case means $F << \frac{(B-1)k_BT}{x_0}$. The range of its validity has been discussed above. With such approximations, Taylor expansion goes on as follows: we introduce $q = \varepsilon B$ and rewrite 10.2 as $q = Z+ln(q)+ln(1+\frac{1}{q})$ where $Z = ln\frac{A}{|\dot\varepsilon|B^2 s}$. Requiring additionally $X \equiv ln\frac{A}{|\dot\varepsilon|B^2} >> lns$ (a natural requirement for $s \cong 1, \frac{A}{|\dot\varepsilon|B^2} >> 1$), from here one obtains the formulae to determine $q = q(X, lns)$ and, correspondingly, $\varepsilon = \varepsilon(X, lns)$ which, when combined with the known definite integrals $\int_0^\infty exp(-x)ln(x)dx = -\gamma$, $\int_0^\infty exp(-x)ln^2xdx = \frac{\pi^2}{6} + \gamma^2$ [89] finally furnish the Garg results presented in the text: $\langle\varepsilon\rangle = (X + lnX + \gamma)/B$, $\langle\varepsilon^2\rangle = \langle\varepsilon\rangle^2 + \frac{\pi^2}{6B^2}$.

Now let us consider the case of positive β. Here we simply take $B = \tilde B$ and get, instead of 10.1, $p(\varepsilon) = \lambda(\varepsilon)\frac{1}{|\dot\varepsilon|}exp(-\int_1^\varepsilon \frac{\lambda(\varepsilon')}{|\dot\varepsilon'|}d\varepsilon')$ and

$$s = \frac{A}{|\dot\varepsilon|} \int_1^\varepsilon \varepsilon' exp(B\varepsilon')d\varepsilon' = \frac{A}{|\dot\varepsilon|B^2}(B\varepsilon exp(\varepsilon B) - exp(\varepsilon B) - Bexp(B) + exp(B))$$

(10.3)

We again require $B\varepsilon >> 1$, and now $\varepsilon = 1 + \beta F > 1$. We will consider only the case when the terms $-Bexp(B) + exp(B)$ can be neglected. This is true whenever $F >> k_BT/x_0$ which is, of course, exactly our region of interest. For such a situation, we can approximate 10.3 with $s = \frac{A}{|\dot\varepsilon|B^2}exp(B\varepsilon)(B\varepsilon - 1)$ and thus

$$lns = -ln\frac{|\dot\varepsilon|B^2}{A} + B\varepsilon + ln(B\varepsilon) + ln(1 - \frac{1}{B\epsilon})$$

(10.4)

Contrary to the above case of negative β, the condition $\frac{A}{|\dot\varepsilon|B^2} << 1$ is to be fulfilled now. This means $\dot F >> k_{off}B^{-2}exp(-B)/|\beta|$ which is always true

for all reasonable force loading rates when $\dot{F} > k_{off}k_BT/x_0 \equiv k_{off}/(B|\beta|)$. Rather cumbersome Taylor expansion with the same approximations as above, viz. $X = -ln\frac{A}{|\dot{\varepsilon}|B^2} >> lns$, and along the same lines, gives eq.10.5

$$B\varepsilon = X - lnX + \frac{lnX}{X} + \frac{1}{X} - \frac{1}{2X^2} + \frac{ln^2X}{2X^2} + ln(s)\left(1 - \frac{1}{X} - \frac{lnX}{X^2}\right) + ln^2s\left(\frac{1}{2X^2}\right) + O(\frac{\ln^3 X}{X^3})$$
(10.5)

Thus with the precision of lnlnX we get:
$\langle\varepsilon\rangle = (X - lnX - \gamma)/B$,$\langle\varepsilon\rangle^2 + \frac{\pi^2}{6B^2}$.

10.4 Tables of Backcalculations

Experimental data and backcalculation						
n	6	5	4	3	2	1
Average - experiment	124	128	129	137	146	162
Average - $n_0 = 1$ backcalculation	127	129	132	137	144	162
Average - $n_0 = 2$ backcalculation	123	126	130	136	146	168
Average - $n_0 = 3$ backcalculation	120	124	129	137	149	170
Average - $n_0 = 4$ backcalculation	118	123	129	137	150	170
Average - $n_0 = 5$ backcalculation	123	128	134	142	153	170
Average - $n_0 = 6$ backcalculation	124	128	134	141	151	164
Average - experiment backcalculation	25	30	31	31	30	31
Average - $n_0 = 1$ backcalculation	11	13	15	18	22	31
Average - $n_0 = 2$ backcalculation	22	23	24	25	30	40
Average - $n_0 = 3$ backcalculation	25	26	28	31	35	39
Average - $n_0 = 4$ backcalculation	27	29	31	34	37	38
Average - $n_0 = 5$ backcalculation	28	30	31	33	34	32
Average - $n_0 = 6$ backcalculation	25	27	28	30	30	28

Table 10.1: Average and standard deviations of the unfolding forces as obtained from the experiment and from the backcalculation using data from $n = 1$ to $n = 6$.

10.5 Kernel Density Estimation

The Kernel Density Estimation (KDE) is a non-parametric method to estimate the distribution of probabilities of a certain set of values. The KDE is a function "f(x)" express in eq.10.6.

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h})$$
(10.6)

Where "K" is the kernel of the function and the parameter "h" is called "bandwidth" of the kernel.
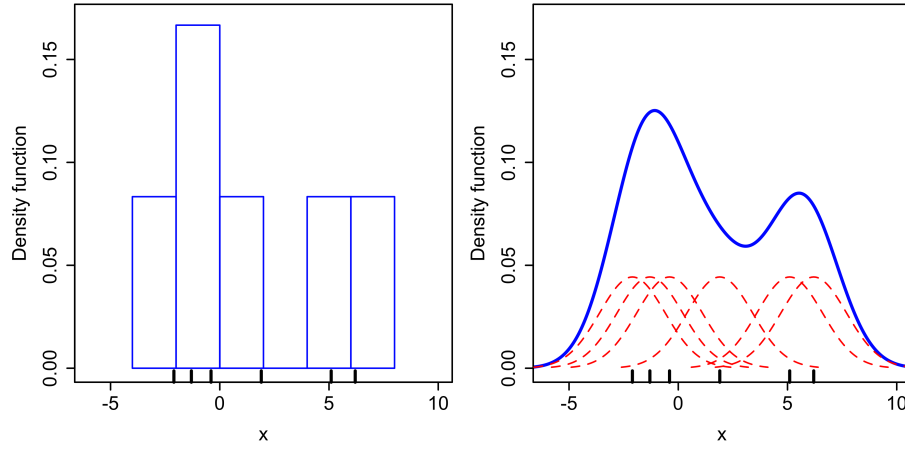
Figure 10.1: In this image we can see the same set of data represented with: (left) an histogram, (right) a gaussian KDE. The red dashed lines are the KDE for each data point, instead the blue is the sum of them.

In particular f(x) has the following properties:

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{10.7}$$

The kernel "K" is often symmetric, however this is not always required. The bandwith "h", in case of gaussian kernel is selected with the "Scott factor" or also so called "Rule of Thumb" expressed in eq. 10.8, where "d" is the dimensionality of the sample, "n" the number of data points and σ_j the standard deviation of the data in such dimension:

$$h_j = n^{-1/(d+4)}\sigma_j \tag{10.8}$$

Usually, to show a probability distribution, histograms are used. Histogram are particular case of the KDE where the kernel is defined as $K(x) = \sharp\left(x\epsilon(a_i, a_{i+1})\right)$ where the range $x_0..x_n$ of data has been partitioned in a_{i+1} sections. If we define $\Delta = a_{i+1} - a_i$, we can define the histogram as:

$$f(x) = \frac{1}{n\Delta}K(x) \tag{10.9}$$

Despite the simple and intiutive procedure to create an histogram this has two problems. The first one is the correct selection of the range,in particular the initial position of the data binning, the second problem is the selection of the size of the partition Δ.

The KDE instead, can overcome these limitations, in fact there is not a division of the range and a selection of the initial position for the binning, moreover the bandwith "h", which is similar to the size of partition Δ has an accurate derivation[90].

Bibliography

[1] F. Valle, M. Favre, P. De Los Rios, A. Rosa, and G. Dietler. Scaling exponents and probability distributions of dna end-to-end distance. *Physical review letters*, 95(15):158105, 2005.

[2] E. Ercolini, F. Valle, J. Adamcik, G. Witz, R. Metzler, P. De Los Rios, J. Roca, and G. Dietler. Fractal dimension and localization of dna knots. *Physical review letters*, 98(5):58102, 2007.

[3] J. Adamcik, D.V. Klinov, G. Witz, S.K. Sekatskii, and G. Dietler. Observation of single-stranded dna on mica and highly oriented pyrolytic graphite by atomic force microscopy. *FEBS letters*, 580(24):5671–5675, 2006.

[4] E. Westhof and P. Auffinger. Rna tertiary structure. *Encyclopedia of analytical chemistry*, 2000.

[5] A.L. Fink. Natively unfolded proteins. *Current opinion in structural biology*, 15(1):35–41, 2005.

[6] T.N. Petersen, C. Lundegaard, M. Nielsen, H. Bohr, J. Bohr, S. Brunak, G.P. Gippert, and O. Lund. Prediction of protein secondary structure at 80% accuracy. *Proteins: Structure, Function, and Bioinformatics*, 41(1):17–20, 2000.

[7] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popović, et al. Predicting protein structures with a multiplayer online game. *Nature*, 466(7307):756–760, 2010.

[8] M. Rief, M. Gautel, F. Oesterhelt, J.M. Fernandez, and H.E. Gaub. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276(5315):1109, 1997.

[9] Y. Cao and H. Li. Polyprotein of GB1 is an ideal artificial elastomeric protein. *Nature Materials*, 6(2):109–114, 2007.

[10] A. Minajeva, M. Kulke, J.M. Fernandez, and W.A. Linke. Unfolding of titin domains explains the viscoelastic behavior of skeletal myofibrils. *Biophysical Journal*, 80(3):1442–1451, 2001.

[11] S. Lv, D.M. Dudek, Y. Cao, MM Balamurali, J. Gosline, and H. Li. Designed biomaterials to mimic the mechanical properties of muscles. *Nature*, 465(7294):69–73, 2010.

[12] G. Meyer and N.M. Amer. Novel optical approach to atomic force microscopy. *Applied Physics Letters*, 53(12):1045–1047, 1988.

[13] E.L. Florin, M. Rief, H. Lehmann, M. Ludwig, C. Dornmair, VT Moy, and HE Gaub. Sensing specific molecular interactions with the atomic force microscope. *Biosensors and Bioelectronics*, 10(9-10):895–901, 1995.

[14] B. Ohler. Practical advice on the determination of cantilever spring constants. *Spring*, pages 1–12, 2007.

[15] M. Schlierf, F. Berkemeier, and M. Rief. Direct observation of active protein folding using lock-in force spectroscopy. *Biophysical journal*, 93(11):3989–3998, 2007.

[16] M. Rief, H. Clausen-Schaumann, H.E. Gaub, et al. Sequence-dependent mechanics of single dna molecules. *nature structural biology*, 6:346–350, 1999.

[17] W. Kuhn and F. Grun. Relationships between elastic constants and stretching double refraction of highly elastic substances. *Kolloid z*, 101:248, 1942.

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

[19] H. Li, M. Rief, F. Oesterhelt, H.E. Gaub, X. Zhang, and J. Shen. Single-molecule force spectroscopy on polysaccharides by afm-nanomechanical fingerprint of [alpha]-(1, 4)-linked polysaccharides. *Chemical physics letters*, 305(3-4):197–201, 1999.

[20] HA Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

[21] G.I. Bell. Models for the specific adhesion of cells to cells. *Science*, 200(4342):618, 1978.

[22] E. Evans and K. Ritchie. Dynamic strength of molecular adhesion bonds. *Biophysical Journal*, 72(4):1541–1555, 1997.

[23] D.J. Brockwell, G.S. Beddard, J. Clarkson, R.C. Zinober, A.W. Blake, J. Trinick, P.D. Olmsted, D.A. Smith, and S.E. Radford. The effect of core destabilization on the mechanical resistance of i27. *Biophysical journal*, 83(1):458–472, 2002.

[24] G. Hummer and A. Szabo. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophysical journal*, 85(1):5–15, 2003.

[25] A.F. Oberhauser, C. Badilla-Fernandez, M. Carrion-Vazquez, and J.M. Fernandez. The mechanical hierarchies of fibronectin observed with single-molecule afm. *Journal of molecular biology*, 319(2):433–447, 2002.

[26] A.F. Oberhauser, P.E. Marszalek, H.P. Erickson, J.M. Fernandez, et al. The molecular elasticity of the extracellular matrix protein tenascin. *Nature*, 393(6681):181–184, 1998.

[27] A.M. Gronenborn, D.R. Filpula, N.Z. Essig, A. Achari, M. Whitlow, P.T. Wingfield, and GM Clore. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein g. *Science*, 253(5020):657–661, 1991.

[28] M. Carrion-Vazquez, A.F. Oberhauser, S.B. Fowler, P.E. Marszalek, S.E. Broedel, J. Clarke, and J.M. Fernandez. Mechanical and chemical unfolding of a single protein: A comparison. *Proceedings of the National Academy of Sciences*, 96(7):3694–3699, 1999.

[29] C. Friedsam, A.K. Wehle, F. Kuhner, and H.E. Gaub. Dynamic single-molecule force spectroscopy: bond rupture analysis with variable spacer length. *JOURNAL OF PHYSICS CONDENSED MATTER*, 15(18):1709–1724, 2003.

[30] D.J. Brockwell, G.S. Beddard, E. Paci, D.K. West, P.D. Olmsted, D.A. Smith, and S.E. Radford. Mechanically unfolding the small, topologically simple protein L. *Biophysical journal*, 89(1):506–519, 2005.

[31] G. Yoon, H.J. Park, S. Na, and K. Eom. Mesoscopic model for mechanical characterization of biological protein materials. *Journal of Computational Chemistry*, 2008.

[32] D.B. Staple, S.H. Payne, A.L.C. Reddin, and H.J. Kreuzer. Stretching and unfolding of multidomain biopolymers: a statistical mechanics theory of titin. *Physical Biology*, 6:025005, 2009.

[33] R. Rounsevell, J.R. Forman, and J. Clarke. Atomic force microscopy: mechanical unfolding of proteins. *Methods*, 34(1):100–111, 2004.

[34] H. Dietz and M. Rief. Elastic bond network model for protein unfolding mechanics. *Physical review letters*, 100(9):98101, 2008.

[35] Y. Cao, C. Lam, M. Wang, and H. Li. Nonmechanical Protein Can Have Significant Mechanical Stability. *Angew. Chem. Int. Ed*, 45:642–645, 2006.

[36] H. Li. Engineering proteins with tailored nanomechanical properties: a single-molecule approach. *Organic & Biomolecular Chemistry*, 5(21):3399–3406, 2007.

[37] M. Sandal, F. Benedetti, M. Brucale, A. Gomez-Casado, and B. Samorì. Hooke: an open software platform for force spectroscopy. *Bioinformatics*, 25(11):1428, 2009.

[38] BW Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, 1986.

[39] B. Efron and R.J. Tibshirani. *An introduction to the bootstrap*. Chapman & Hall, 1997.

[40] M. Schlierf and M. Rief. Single-molecule unfolding force distributions reveal a funnel-shaped energy landscape. *Biophysical journal*, 90(4):L33–L35, 2006.

[41] R.C. Zinober, D.J. Brockwell, G.S. Beddard, A.W. Blake, P.D. Olmsted, S.E. Radford, and D.A. Smith. Mechanically unfolding proteins: The effect of unfolding history and the supramolecular scaffold, 2002.

[42] R. Berkovich, S. Garcia-Manyes, M. Urbakh, J. Klafter, and J.M. Fernandez. Collapse dynamics of single proteins extended by force. *Biophysical journal*, 98(11):2692, 2010.

[43] S. Sekatskii F. Benedetti and G. Dietler. Single-molecule force spectroscopy of multimodular proteins: a new method to extract kinetic unfolding parameters. *Journal of Scanning Probe Microscopy*, In press-2011.

[44] E. Barkai and M. Orrit. *Theory and evaluation of single-molecule signals*. World Scientific Pub Co Inc, 2008.

[45] E.J. Gumbel. *Statistics of extremes*. Dover Pubns, 2004.

[46] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, et al. *Numerical recipes*, volume 3. Cambridge university press Cambridge, 2007.

[47] Á. Karsai, M.S.Z. Kellermayer, and S.P. Harris. Mechanical unfolding of cardiac myosin binding protein-c by atomic force microscopy. *Biophysical journal*, 101(8):1968–1977, 2011.

[48] C. Bouchiat, MD Wang, J.F. Allemand, T. Strick, SM Block, and V. Croquette. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophysical journal*, 76(1):409–413, 1999.

[49] C. Ke, Y. Jiang, M. Rivera, R.L. Clark, and P.E. Marszalek. Pulling geometry-induced errors in single molecule force spectroscopy measurements. *Biophysical journal*, 92(9):L76–L78, 2007.

[50] Z. Lu, W. Nowak, G. Lee, P.E. Marszalek, and W. Yang. Elastic properties of single amylose chains in water: A quantum mechanical and afm study. *Journal of the American Chemical Society*, 126(29):9033–9041, 2004.

[51] P.E. Marszalek, H. Li, J.M. Fernandez, et al. Fingerprinting polysaccharides with single-molecule atomic force microscopy. *Nature biotechnology*, 19(3):258–262, 2001.

[52] M. Rief, J.M. Fernandez, and H.E. Gaub. Elastically Coupled Two-Level Systems as a Model for Biopolymer Extensibility. *Physical Review Letters*, 81(21):4764–4767, 1998.

[53] W.T. King, M. Su, and G. Yang. Monte carlo simulation of mechanical unfolding of proteins based on a simple two-state model. *International journal of biological macromolecules*, 46(2):159–166, 2010.

[54] M. Rief, F. Oesterhelt, B. Heymann, and H.E. Gaub. Single molecule force spectroscopy on polysaccharides by atomic force microscopy. *Science*, 275(5304):1295, 1997.

[55] S. Allen, X. Chen, J. Davies, M.C. Davies, A.C. Dawkes, J.C. Edwards, C.J. Roberts, J. Sefton, S.J.B. Tendler, and P.M. Williams. Detection of antigen-antibody binding events with the atomic force microscope. *Biochemistry*, 36(24):7457–7463, 1997.

[56] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophysical journal*, 72(4):1568–1581, 1997.

[57] M. DE ODROWAZ PIRAMOWICZ, P. CZUBA, M. TARGOSZ, K. BURDA, and M. SZYMONSKI. Dynamic force measurements of avidin-biotin and streptavdin-biotin interactions using afm. *Acta Biochimica Polonica*, 53(1):93–100, 2006.

[58] V.T. Moy, E.L. Florin, and H.E. Gaub. Adhesive forces between ligand and receptor measured by afm. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 93:343–348, 1994.

[59] J. Zlatanova, S.M. Lindsay, and S.H. Leuba. Single molecule force spectroscopy in biology using the atomic force microscope. *Progress in biophysics and molecular biology*, 74(1-2):37–61, 2000.

[60] F. Schwesinger, R. Ros, T. Strunz, D. Anselmetti, H.J. Güntherodt, A. Honegger, L. Jermutus, L. Tiefenauer, and A. Plückthun. Unbinding forces of single antibody-antigen complexes correlate with their thermal dissociation rates. *Proceedings of the National Academy of Sciences*, 97(18):9972, 2000.

[61] J.W. Weisel, H. Shuman, and R.I. Litvinov. Protein-protein unbinding induced by force: single-molecule studies. *Current opinion in structural biology*, 13(2):227–235, 2003.

[62] C.K. Lee, Y.M. Wang, L.S. Huang, and S. Lin. Atomic force microscopy: determination of unbinding force, off rate and energy barrier for protein-ligand interaction. *Micron*, 38(5):446–461, 2007.

[63] O.K. Dudko, G. Hummer, and A. Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Physical review letters*, 96(10):108101, 2006.

[64] P. Hänggi, P. Talkner, and M. Borkovec. Reaction-rate theory: fifty years after kramers. *Reviews of Modern Physics*, 62(2):251, 1990.

[65] 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*, 105(41):15755, 2008.

[66] Y. Suzuki and O.K. Dudko. Single-molecule rupture dynamics on multidimensional landscapes. *Physical review letters*, 104(4):48101, 2010.

[67] A. Garg. Escape-field distribution for escape from a metastable potential well subject to a steadily increasing bias field. *Physical Review B*, 51(21):15592–15595, 1995.

[68] J. Kurkijärvi. Intrinsic fluctuations in a superconducting ring closed with a josephson junction. *Physical Review B*, 6(3):832, 1972.

[69] M. Favre, LA Chtcheglova, DA Lapshin, SK Sekatskii, F. Valle, and G. Dietler. Force-clamp spectroscopy with a small dithering of afm tip, and its application to explore the energy landscape of single avidin-biotin complex. *Ultramicroscopy*, 107(10-11):882–886, 2007.

[70] F. Pincet and J. Husson. The solution to the streptavidin-biotin paradox: the influence of history on the strength of single molecular bonds. *Biophysical journal*, 89(6):4374–4381, 2005.

[71] R. Merkel, P. Nassoy, A. Leung, K. Ritchie, and E. Evans. Energy landscapes of receptor–ligand bonds explored with dynamic force spectroscopy. *Nature*, 397(6714):50–53, 1999.

[72] F. Benedetti, C. Micheletti, G. Bussi, SK Sekatskii, and G. Dietler. Nonkinetic modeling of the mechanical unfolding of multimodular proteins: Theory and experiments. *Biophysical Journal*, 101(6):1504–1512, 2011.

[73] J.F. Marko and E.D. Siggia. Stretching dna. *Macromolecules*, 28(26):8759–8770, 1995.

[74] AV Sorokin, AM Mikhailov, AV Kachko, EV Protopopova, SN Konovalova, ME Andrianova, SV Netesov, AN Kornev, VB Loktev, et al. Human recombinant laminin-binding protein: isolation, purification, and crystallization. *BIOCHEMISTRY C/C OF BIOKHIMIIA*, 65(5):546–553, 2000.

[75] G. Martin and R.K. Jain. Noninvasive measurement of interstitial ph profiles in normal and neoplastic tissue using fluorescence ratio imaging microscopy. *Cancer research*, 54(21):5670, 1994.

[76] JR Griffiths. Are cancer cells acidic? *British journal of cancer*, 64(3):425, 1991.

[77] M.S.Z. Kellermayer, L. Grama, Á. Karsai, A. Nagy, A. Kahn, Z.L. Datki, and B. Penke. Reversible mechanical unzipping of amyloid β-fibrils. *Journal of Biological Chemistry*, 280(9):8464–8470, 2005.

[78] M. Kim, C.C. Wang, F. Benedetti, and P.E. Marszalek. A nanoscale force probe for gauging intermolecular interactions. *Angewandte Chemie*, 124(8):1939–1942, 2012.

[79] M. Kim, C.C. Wang, F. Benedetti, M. Rabbi, V. Bennett, and P.E. Marszalek. Nanomechanics of streptavidin hubs for molecular materials. *Advanced Materials*, 2011.

[80] M. Sandal, F. Valle, I. Tessari, S. Mammi, E. Bergantino, F. Musiani, M. Brucale, L. Bubacco, and B. Samorì. Conformational Equilibria in Monomeric α-Synuclein at the Single-Molecule Level. *PLoS Biol*, 6(1):e6, 2008.

[81] C.C. Wang, T.Y. Tsong, Y.H. Hsu, and P.E. Marszalek. Inhibitor binding increases the mechanical stability of staphylococcal nuclease. *Biophysical Journal*, 100(4):1094, 2011.

[82] A. Pähler, WA Hendrickson, MA Kolks, CE Argarana, and CR Cantor. Characterization and crystallization of core streptavidin. *Journal of Biological Chemistry*, 262(29):13933–13937, 1987.

[83] P.C. Weber, DH Ohlendorf, JJ Wendoloski, and FR Salemme. Structural origins of high-affinity biotin binding to streptavidin. *Science*, 243(4887):85–88, 1989.

[84] W.A. Hendrickson, A. Pähler, J.L. Smith, Y. Satow, E.A. Merritt, and R.P. Phizackerley. Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proceedings of the National Academy of Sciences*, 86(7):2190, 1989.

[85] I.P. Korndörfer and A. Skerra. Improved affinity of engineered streptavidin for the strep-tag ii peptide is due to a fixed open conformation of the lid-like loop at the binding site. *Protein science*, 11(4):883–893, 2002.

[86] K.A. Dickson, M.C. Haigis, and R.T. Raines. Ribonuclease inhibitor: structure and function. *Progress in nucleic acid research and molecular biology*, 80:349–374, 2005.

[87] L.A. Chtcheglova, G.T. Shubeita, S.K. Sekatskii, and G. Dietler. Force spectroscopy with a small dithering of afm tip: a method of direct and continuous measurement of the spring constant of single molecules and molecular complexes. *Biophysical journal*, 86(2):1177–1184, 2004.

[88] NA Burnham, AJ Kulik, G. Gremaud, P.J. Gallo, and F. Oulevey. Scanning local-acceleration microscopy. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, 14(2):794–799, 1996.

[89] Gradshteĭ. *Table of integrals, series, and products.*

[90] W. Härdle, A. Werwatz, M. Müller, and S. Sperlich. Introduction. *Nonparametric and semiparametric models*, pages 1–18, 2004.

Fabrizio BENEDETTI

Address:
Rue De la Mouline 6, 1022
Chavannes-près-renens,
Switzerland

Personal informations:
12/06/1982
Single
Italian
fabrizio.benedetti.82@gmail.com
fabrizio.benedetti@epfl.ch
Mobile: "+41798647216"

SUMMARY

Motivated science graduate. Interested in modeling of physical systems, research in physics and signal processing. Would like to widening his knowledge and find a challenging position to meet his competencies, capabilities and skills. Good in problem solving.

WORK EXPERIENCE

PhD Student   01/08/08 to present
École Polytechnique Fédérale de Lausanne (EPFL, Lausanne, Switzerland)
- Teaching assistant
- Research in biophysics
- Experimental study of the mechanical properties of the proteins and of organic chains
- Signal processing and data analysis
- Chemical functionalization of surfaces
- Develop of algorithm for signal processing
- Monte Carlo simulations of elastic chains with bonds breaking

Invited PhD student   11/01/11 to 30/4/11
Duke University (Durham, NC, USA)
- Experiment of protein pulling
- Biochemical synthesis of multimodular proteins
- Data analysis and statistics

Trainee   01/05/00 to 01/06/00
Valli Zabban SPA (Monterado, Italy)
- Quality control of the production with physical and chemical tests

EDUCATION

PhD student in Biophysics   01/08/08, estimated graduation 08/12
École Polytechnique Fédérale de Lausanne (EPFL, Lausanne, Switzerland)

Master of Science in Applied Physics   Graduation 20/06/08
Alma Mater Studiorum (Università degli studi di Bologna)
Bachelor in Physics   Graduation 14/10/05
Alma Mater Studiorum (Università degli studi di Bologna)

Secondary school diploma in industrial chemical technique   2001
Istituto tecnico industriale "Vito Volterra" (Torrette, Italy)

LANGUAGES

Italian: mother tongue
English: good knowledge (B2 level certificate)
French: intermediate (A2 level certificate)

COMPUTER SKILLS

Operating systems: Windows (Good), Linux (Good), MacOS (poor).
Programming languages: Python (Good), C++ (Good), LaTeX (Good), Java (elements), Matlab (elements), Labview (elements), CUDA(elements).
Suite: MS-Office, OpenOffice.

LIST OF PUBLICATIONS

Hooke: an open software platform for force spectroscopy
Sandal, M. and Benedetti, F. and Brucale, M. and Gomez-Casado, A. and Samori, B.
Bioinformatics,
volume 25, number 11, pages 1428, year 2009

Single-Molecule Force Spectroscopy of Multimodular Proteins: A New Method to Extract Kinetic Unfolding Parameters
Benedetti, F. and Sekatskii, S.K. and Dietler, G.
Journal of Advanced Microscopy Research,
volume 6, number 1, pages 1—6, year 2011

Nonkinetic Modeling of the Mechanical Unfolding of Multimodular Proteins: Theory and Experiments
F. Benedetti, C. Micheletti, G. Bussi, S.K. Sekatskii and G. Dietler
Biophysical Journal,
volume 101, number 6, pages 1504—1512, year 2011

Nanomechanics of Streptavidin Hubs for Molecular Materials
Kim, M. and Wang, C.C. and Benedetti, F. and Rabbi, M. and Bennett, V. and Marszalek, P.E.
Advanced Materials,
Volume 23, Issue 47, pages 5684-6688, year 2011

A Nanoscale force probe for gauging intermolecular interactions
Kim M., Wang, C.C. and Benedetti, F. and Marszalek, P.E.
Angewandte Chemie International Edition,
volume 124, issue 8, pages 1939-1942
DOI: 10.1002/anie.201107210, year 2012

Dependence of the most probable and average bond rupture force on the force loading rate: first order correction to the Bell – Evans model
S. K. Sekatskii, F. Benedetti, and G. Dietler
(Article in preparation)

An extraction of dissipative properties of single receptor – ligand complex from the phase data obtained exploring a small-amplitude dithering of AFM tip
S. K. Sekatskii, F. Benedetti, A. Kulik, P.E. Marszalek and G. Dietler
(Article in preparation)

OTHER

Wide experience in physical and chemical laboratory.
Italian driving license B (car).
Other work experiences far from the study.