Università degli Studi di Milano-Bicocca

DIPARTIMENTO DI SCIENZA DEI MATERIALI
Corso di Dottorato in Scienza e Nanotecnologia dei Materiali
XXXII ciclo

Tesi di Dottorato

Single Molecule Force Spectroscopy of Proteins and DNA

Candidato:
Roberta Corti
736505

Tutor:
Dott. Domenico Salerno

Co-Tutor:
Dott.ssa Valeria Cassina

Coordinatore:
Prof. Marco Bernasconi

Anno Accademico 2018-2019

CONTENTS

INTRODUCTION

In the last few decades, the constant development of novel single molecule techniques has created the basis for new paradigms in the field of biophysics. Among all, the nanomanipulation of individual biomolecules revealed relevant insights into the mechanics of biological molecules, in particular proteins and DNA, improving the understanding of the fundamental relation between structural properties and biological functions. Therefore, several single-molecule nanomanipulation methods have been developed, including Atomic Force Microscopy (AFM), Magnetic Tweezers (MT) and Flow Stretching (F-S) coupled with fluorescence. All these technique were employed in this Thesis for the characterisation of biological macromolecules by Single Molecule Force Spectroscopy (SMFS). Specifically, an AFM and a MT setup, both located in the biomedical and biophysics laboratory of Prof. Francesco Mantegazza, at the University of Milano-Bicocca, were used in the first two projects, while another MT and a F-S coupled with TIRF were employed at the Single-Molecule DNA-repair nanomachines laboratory of Dott. Fernando Moreno-Herrero, at the Spanish National Center of Biotechnology in Madrid.
In this Thesis I focus mainly on several aspects of few different proteins trying to depict a frame in which the strong link between proteins function and structure can be clarified. With this aim, I study the conformational states of an intrinsically disordered protein (IDP) involved in Parkinson's Disease (PD), the α-synuclein, and I analyse the structural change driving the DNA compaction mediated by a structural maintenance protein, the condensin. Finally, I present a structural study of a diamino-substituted DNA-analogue by using thermal shifting assays and single molecule experiments. I include also a technical implementation of a F-S combined with TIRF set up to promote the high-speed exchange of buffer to study protein:DNA interactions.

This thesis is divided into 6 main parts: the *State of Art*, the 4 projects and a final *Summary*. Each of those projects is, in its turn, divided into 4 chapter: the *Introduction*, where the project is briefly presented, the *Materials and Methods*, in which the details about the investigated biological samples, the experimental procedures and the data analysis are described. Then, in the *Results* chapters the main findings are reported and, finally, in the *Discussion* part the results are discussed and compared to the literature.

- In the first part of this thesis, *State of Art*, I present all the single molecule techniques used in this work, both theoretically

and experimentally as well as the biological macromolecules under investigation. Firstly, the AFM setup is described, from the operation principles, to the calibration procedures necessary to perform the typical measurements. In particular, I focus on the force spectroscopy mode, which was widely employed to characterized the conformational structures of the AS protein. I briefly explain the imaging mode as well, used during the DNA analogue studies at zero force. Then, I present the MT setup employed for both DNA-analogue studies and condensin mediated DNA condensation. Finally, I describe the TIRF-based flow stretching apparatus used in Madrid for the visualisation of stretched DNA. In the Chapter 2, I briefly discuss the biology of proteins, focusing on their structure and function. I present a concise review on the characteristics of intrinsically disordered proteins and, in particular, the α-synuclein (AS) protein. I concentrate on the AS structure and its role in Parkinson disease insurgence, including its genetic and pathological point mutations. I present also the recent finding of structural rearrangements of the AS as a response to the presence of external ligands. Finally, in the Chapter 3 I briefly present the biological structure of the DNA and its interactions with structural maintenance chromosome complexes. I mainly focus on the models of DNA compaction driven by the condensin complex.

- In the *Depicting Conformational Ensembles of α-Synuclein by Single Molecule Force Spectroscopy* study, presented in the part II, I afford the problem of AS structure classification by stretching and unfolding a single polyprotein containing the human AS by employing a SMFS approach. The analysis of the different unfolding pathways gives information about the structural conformation of the protein before the mechanical denaturation. The AS was found to assume three distinct conformational states ranging from a random coil to a highly structured conformation. Since ligands, such as Epigallocatechin-3-Gallate (EGCG) and Dopamine (DA), are known to affect the fibrillation process of AS, I used this single molecule technique to investigate the effect of EGCG and DA on the conformational ensemble of WT AS. Moreover, knowing from several studies that the presence of point mutations, linked to familial PD, correlate with the gaining of structure and therefore with AS aggregation, I performed SMFS studies also on AS with three different single point mutations (A30P, A53T and E83A). A particular emphasis was given to the comparison between SMFS results and native mass spectrometry data for the conformational changes of AS in the presence of both DA and EGCG.

- In the following part (III), entitled *Nanomechanical Characterisation of Diaminopurine-Substituted DNA*, I discuss the nanomechanical characterization of a DNA analogue, namely DAP DNA. A systematic comparison between a wild-type DNA and DAP DNA is performed, in terms of thermal stability and nanomechanical properties, measured at low and high forces. By using both MT and AFM, the DNA extension and bending rigidity were investigated at low forces while at high forces the overstretching transition behavior was explored.

- In the part IV, *Direct observation of Condensin mediated DNA-collapsing*, I present a single-molecule MT study to measure, in real-time, the compaction of individual DNA molecules induced by the condensin complex in the presence of ATP. Since many compaction traces showed sudden distinct decreases in the DNA end-to-end length, I present and validate two different very conservative user-bias-independent step-finding algorithm to extract the size of these compaction steps.

- In the part V, the *DNA flow stretching coupled with single molecule fluorescence* implementation is presented. Briefly, several flow cells were tested to achieve a fast buffer exchange in both MT and F-S coupled with TIRF, in the frame of visualisation of DNA:proteins interactions. We validated our flow cells in term of boundary exchange and applied force. We also visualized fluorescent DNA molecules stretched by the flow.

- Finally, I include a brief *Summary* to summarize all the findings and the results obtained in all the projects presented here.

In addition, since, during my PhD, I was involved into several works which are not directly connected with the force spectroscopy of biomacromolecules, I include some of them in the Appendix A and B.

LIST OF PUBLICATIONS

1. **R. Corti**, C. A. Marrano, D. Salerno, S. Brocca, A. Natalello, C. Santambrogio, G. Legname, F. Mantegazza, R. Grandori and V. Cassina. *Depicting confoormational ensembles of a-synuclein by single molecule force spectroscopy and native mass spectroscopy*, International Journal of Molecular Sciences, **20**, (20), 5181, (2019).

2. M. Cristofalo, D. Kovari, **R. Corti**, D. Salerno, V. Cassina, D. Dunlap, F. Mantegazza. *Nanomechanics of Diaminopurine-Substituted DNA*, Biophysical Journal, **116**, 760-771, (2019).

3. I. Figuereido, A. Paiotta, R. Dal Magro, F. Tinelli, **R. Corti**, F. Re, V. Cassina, E. Caneva, F. Nicotra, L. Russo. *A New Approach for Glyco-Functionalization of Collagen-Based Biomaterials*, International Journal of Molecular Sciences, **20**, 1747, (2019).

4. R. Dal Magro, S. Simonelli, A. Cox, B. Formicola, **R. Corti**, V. Cassina, L. Nardo, F. Mantegazza, D. Salerno, G. Grasso, M. A. Deriu, A. Danani, L. Calabresi, F. Re. *The extent of human apolipoprotein A-I lipidation strongly affects the beta amyloid efflux across the blood-brain barrier in vitro*, Frontiers in Neuroscience, **13**, 419, (2019).

5. B. Torsello, S. De Marco, S. Bombelli, E. Chisci, V. Cassina, **R. Corti**, D. Bernasconi, R. Giovannoni, C. Bianchi, R. A. Perego. *The 1ALCTL and 1BLCTL isoforms of Arg/Abl2 induce fibroblast activation and extra cellular matrix remodelling differently*, Biology Open, **8**, bio038554, (2019).

6. V. Di Stefano, B. Torsello, C. Bianchi, I. Cifola, E. Mangano, G. Bovo, V. Cassina, S. De Marco, **R. Corti**, C. Meregalli, S. Bombelli, P. Vigano, C. Battaglia, G. Strada, R. A. Perego. *Major Action of Endogenous Lysyl Oxidase in Clear Cell Renal Cell Carcinoma Progression and Collagen Stiffness Revealed by Primary Cell Cultures*, The American Journal of Pathology, **9**, Vol. 186, (2016).

UNDER REVIEW

7. D. Fontana, M. Mauri, R.Renso, M. Docci, I. Crespiatico, L. Røst, A. Niro, D. D'Aliberti, L. Massimino, M. Bertagna, G. Zambrotta, M. Bossi, S. Citterio, B. Crescenzi, F. Fanelli, V. Cassina, **R. Corti**, D. Salerno, L. Nardo, F. Mantegazza, C. Mecucci,

G. Cavaletti, P. Bruheim, D. Rea, S. Larsen, C. Gambacorti-Passerini, R. Piazza. *ETNK1 mutations induce a mutator phenotype that can be reverted with phosphoethanolamine*, Nature Communication, under review, (2019).

UNDER PREPARATION

8. **R. Corti**, A. Cox , V. Cassina, L. Nardo, D. Salerno, C. A. Marrano, N. Missana, P. Andreozzi, P. J. Silva, F. Stellacci, R. Dal Magro, F. Re, F. Mantegazza. *Clustering on gold nanoparticle surface improves the anti-amyloidogenic effect of mApoE peptide*, under preparation.

9. M. Cristofalo, C. A. Marrano, D. Salerno, **R. Corti**, V. Cassina, B. Sclavi, A. Mammola, M. Gherardi, M. Cosentino Lagomarsino, F. Mantegazza. *Cooperative effects on the compaction of short DNA fragments by the nucleoid protein H-NS and the crowding agent PEG probed by Magnetic Tweezers*, under preparation.

Part I

STATE OF ART

1

BIOPHYSICS AND SINGLE MOLECULE

During the last tens years, the development of several optical techniques combined with biology have completely changed the way to approach the characterisation of the living world. The wide use of advanced microscopy, often supported by a complementary fluorescence approach, has exponentially accelerated the understanding of a great number of biologically relevant questions. In this perspective, it has been essential a more strong interconnection among chemistry, biology, physics and nanotechnology. The biophysics plays a prominent role in this new field, using a physical approach applied to biological systems and biochemical reactions to reveal the forces and energies that drive the biological phenomena. Understanding the living processes is very appealing for a biophysicist, both in the perspective of a pure knowledge and of a new way to treat and prevent several diseases.

1.1 SINGLE MOLECULE

The development of several single molecule techniques have become more and more popular among different fields: biology, biophysics, chemistry, nanoscience and materials characterisation. These techniques have the peculiar feature to focus on a single molecule, avoiding all the problems arising from the use of bulk techniques.

As a matter of facts, even if the atomic model of matter has been accepted since 200 years ago, until few tens of years ago the molecular processes were still hardly accessible. In effect, the molecular scale is too little to be investigated by standard techniques and too big for being characterized by X-ray crystallography or other spectroscopy techniques: molecules have been purified, broken and modified for years, but nobody could analyse them directly and their properties are extrapolated from a bulk of billions of average molecules. This results in a loss of informations about properties which deviate too far from the average: it has been impossible to have punctual analysis of a single and specific molecule. In particular, forces involved in the majority of the biological molecular scale processes, from cellular motility to protein folding, are extremely difficult to be detected by using bulk techniques.

This problem has been overcame in the last 30 years by the development of several single molecule (SM) techniques, including Optical Tweezers (OT), Magnetic Tweezers (MT), Atomic Force Microscopy (AFM) and Flow Stretching (F-S) as described in [10, 11, 12, 13]. The power and resolution of these techniques are highlighted by the wide variety of systems investigated: generally, the capability of single molecules manipulation cover six orders of magnitude in both length ($10^{-10} - 10^{-4}$m) and force ($10^{-14} - 10^{-8}$N). The time scales depends on the great variety of the studied process, but usually are in the range of $10^{-3} - 10^{2}$s. Consequently, the energies involved are considerably lower than the ones of the macroscale, being in the scale of the thermal fluctuations ($k_B T \sim 10^{-21}$J)[1].

The typical SM force range permits the investigation of the elasticity of biochemical polymers, as DNA or RNA, as well as it could provide the measure of bond energies, lifetime and energy landscape of a big variety of proteins [14, 15, 16, 17, 18, 19]. Moreover, the SM techniques are capable of monitoring changes in extensions and forces in the relevant scale with ms resolution, being the perfect tool to address measurements on proteins unfolding and DNA:protein interactions.

One of the most important aspects of single-molecule techniques is that they do not suffer from the main problem which is typical of ensemble measurement: rare phenomena, which in bulk techniques can easily hidden by averaging, could be observed and analysed separately. This is particularly important in the study of heterogeneous ensembles, such as the intrinsically disordered proteins families, or to characterize the interactions of proteins with single DNA filaments. By using a single molecule approach, the molecule itself could be the real object of study and its properties could be monitored quantitatively as specific values, not averaging from millions of molecules as happens in the standard bulk biophysics techniques.

In this Thesis three of this force spectroscopy techniques have been used, probing the versatility of the SM approach to unravel different biological processes at the single molecule level. In particular, (1) the AFM was used to probe the conformational changes of an intrinsically disordered protein, the α-synuclein, in the presence of different ligand or pathological point mutations, (2) the MT was employed to probe the nanomechanics of a DNA analogue and to characterize the interactions between DNA and condensin and (3) the instrumental implementation of F-S coupled with fluorescence for a rapid boundary exchange was validate in order to be used to investigate different protein binding to DNA in future.

The AFM is often used as an imaging characterisation tool, in which the investigation of biological process under zero force could be achieved (as described in 11.1).

1 $k_B = 1.38064910^{-23} \times JK^{-1}$ is the Boltzmann constant

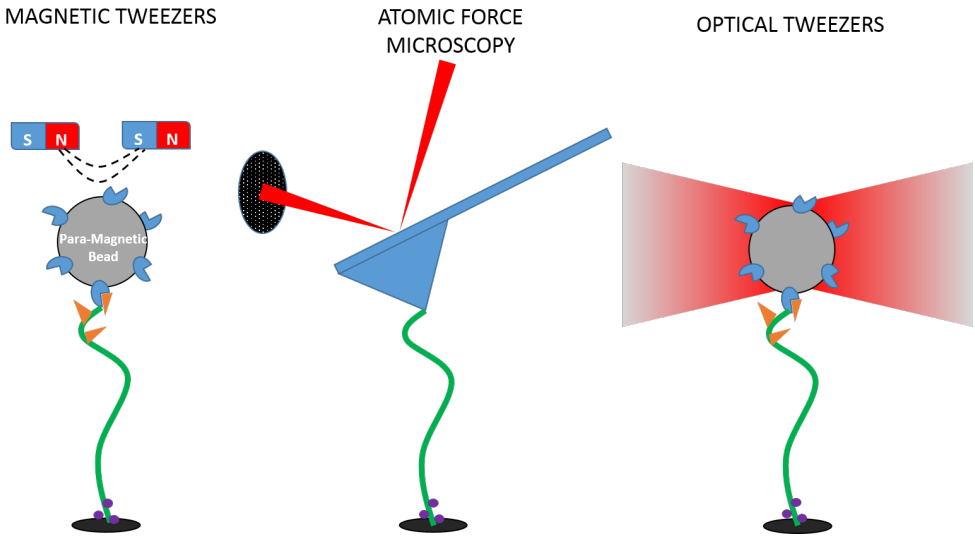

Figure 1.1.: Main single molecule force spectroscopy techniques: Magnetic Tweezers, Atomic Force Microscopy and Optical Tweezers tethering a single DNA filament.

	OT	MT	AFM
Spatial resolution (nm)	$0.1-2$	$5-10$	$0.5-1$
Temporal resolution (s)	10^{-4}	10^{-1}-10^{-2}	10^{-3}
Stiffness (pN $\cdot$ nm^{-1})	$0.005-1$	10^{-3}-10^{-6}	10-10^{5}
Force Range (pN)	$0.1-100$	10^{-3}-10^{2}	10-10^{4}
Displacement Range (nm)	$0.1-10^{5}$	$5-10^{4}$	$0.5-10^{4}$
Probe Size (μm)	$0.25-5$	$0.5-5$	$100-250$
Typical applications	3D manipulation Tethered assay Interaction assay	Tethered assay DNA topology	High Force Pulling Interaction assay
Features	Low noise and low-drift geometry	Force clamp Bead rotation Specific interactions	High resolution imaging Interaction assay
Limitations	Photo damage Sample heating Nonspecific	No manipulations	Large high stiffness probe large minimal force Nonspecific

Table 1.1.: Comparison of the three main force spectroscopy techniques: Optical Tweezers (OT), Magnetic Tweezers (MT) and Atomic Force Microscopy (AFM). Adapted from [10].

1.1.1 *Advantages of Single Molecule approaches*

With *Single Molecule techniques* one refers to several methods which allows the measurements of physical properties on one-to-one molecule [13].

In particular, the information obtained could be divided in:

- **Forces** The force spectroscopy is widely employed into the biophysics field, due to the great range of biological processes in which the force is involved, as mechanical protein unfolding or DNA rearrangements and organisation. The MT and OT can ap-

ply external forces from sub-pN to tens of pN, while the AFM in force spectroscopy mode goes from tens of pN to nN (see Tab.1.1 for details). In the Flow Stretching techniques forces in the order of several pN could be achieved.

- **Structural information** Quite recently, the SM techniques have been used as tools to unravel the structure of biomacromolecules as proteins and nucleic acids, complementary to other more *standard* approach as Mass Spectrography or Crystallography. In the SM, the structural information are obtained by measuring the intermolecular distances, obtained a structural resolution in the order of nm.

- **Dynamics** Usually, all SM approaches allow the investigation of the dynamics of the processes under study, usually with a resolution ranging from ms to several minutes. The folding and unfolding of proteins, or the protein-driven compaction of a DNA are some examples.

In general, among all the advantages, also some drawbacks are present. The unique feature of SM technique to distinguish very rare processes or behaviour could results also in considering spurious events or instrumental-linked artefacts as real phenomenons. Single molecule events are also limited by the thermal noise, which is comparable to the noise of the measure and cannot be avoid. They are also affected by high thermal fluctuations, and sometimes a proper calibration could be challenging. Moreover, to have a reasonable statistic, the measurements are very time-demanding. Despite all this, the SM techniques are necessary to study molecular heterogeneity, the dynamics out of the equilibrium and permit to work with a very low concentration of biological sample.

The first *Atomic Force Microscopy* (AFM) set up was realized in 1986 by Binnig, Quate and Gerber [20] and its principles are based to the *Scanning Probe Microscope* (SPM). In a typical SPM system the value of a physical variable, which depends on the distance between the surface and a specific probe, is monitored. The SPM techniques were born as powerful tools for obtaining 2-dimensional maps of a surface (with several different variables recorded, like interaction force (AFM), electrical current (Scanning Tunneling Microscopy - STM) or small light sources (Near-Field Scanning Optical Microscopy - NSOM)). In the AFM setup, the bending of the cantilever in the vertical axis is the physical variable that is monitored: in the imaging mode the instrument works as a nanoscale phonograph: after having recorded a very high rate of 2-dimensional profiles, the topography of the surface is reconstructed in three-dimensional details (Fig. 1.2).

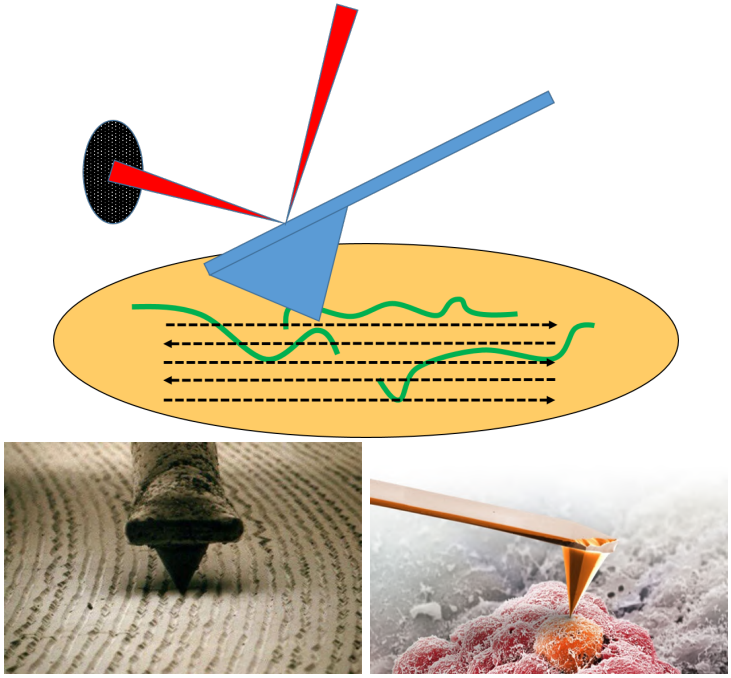

Figure 1.2.: Top: A simplified cartoon of an AFM in imaging mode.
Bottom Left: A phonograph tip on a LP.
Bottom Right: An AFM tip on a biological cell [adapted from RITMindustry®].

Even if this instrument started his history as a powerful tool for nanoscale surface morphology investigation, during the years a lot of different abilities emerged in a wide variety of different fields, from microelectronic to biology to nanotechnology.

At these days, the AFM is the most used instrument for characterising biological molecules as proteins or DNA at nanoscales [2, 4, 8, 9], as well as one-atom-resolution graphene matrices [22] or surfaces as

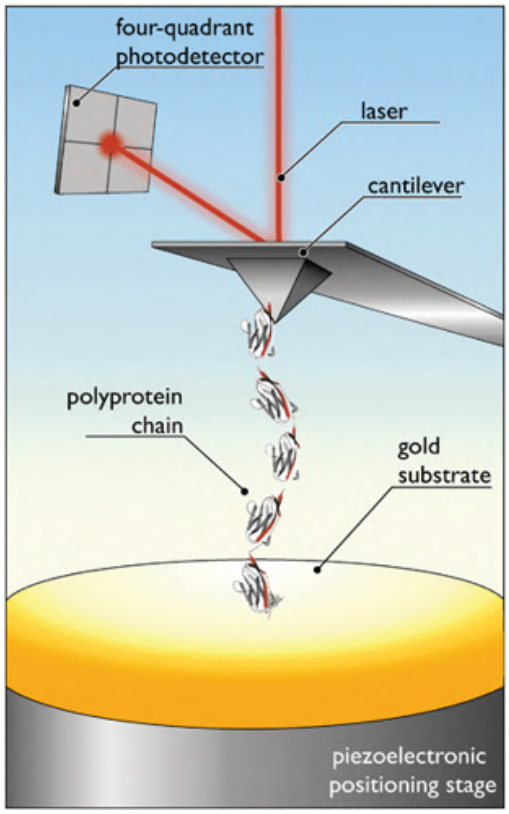

Figure 1.3.: A sketch of the standard setup of an AFM in a protein unfolding experiment. (Fig. taken from [14]).

Figure 1.4.: The AFM Life Science Nanowizard® II from JPK used in this Thesis.

collagen [3]. Important improvements of the setup (High-speed AFM [23], up to 30 frames/s) have allowed to scale from the characterisation of static samples to obtain dynamic information of biological sample under motion, as, for example, proteins [24].

The AFM is nowadays one of the most used tools also for probe the elasticity of single cells [7] or bacteria, or probe the stiffness of samples at the mesoscale (as, for example, collagen matrices [6, 5]).

Indeed, the AFM is also one of the most relevant tools for the nanomanipolation of matter and for probing biological systems properties which are related to the surfaces-tip interaction. It is extremely sensitive to forces, so it has emerged as the main tool for probing the unfolding of proteins [1] or the DNA overstretching at forces up to 20 pN.

The standard setup of a typical AFM (Fig.1.3) consists in a cantilever with a sharp tip at one of the end. A laser beam is reflected to the back of the tip and then recorded by a system of four segmented quadrant photodiode and converted in electrical signal which enters a feedback process. This investigating tip is moved very close to the surface and then retracted by a piezoelectric actuator which records the position with a subnanometer resolution. By analysing the displacement of the laser on the photo diode, lots of different informations could be obtained about the surface or about the interaction between the tip and the sample. The setup is completed by a stage, on which the sample is positioned, an optical camera and, in some version, an optical or fluorescent microscope.

The cantilever usually is made of Silicon Nitride (Si_3N_4) and it could be coated by gold in order to increase the reflectivity of the laser beam. The central body consists on a very tiny slice of Si_3N_4 with a V-shape arms or on a rectangular arm, whose length is in the order of $100\mu m$. The tip is applied on one end of the cantilever. The real tip is very sharp and could have different shapes: for the protein unfolding study it is common to use a quadratic pyramid tip, with a curvature radius of $\sim 10nm$ [1] mounted on a soft cantilever, while for acquiring images a symmetric tip with a nominal tip radius of 8 nm was used [3, 4, 8]. The characteristics of the probe could vary, as the length of the arms on which the tip is mounted or their width: the physical parameters define the final elastic constant of the probe. A very soft cantilever results (as the one usually used for force spectroscopy) in a higher resolution in laser deflection detection: the more the cantilever is bent, the more the laser beam is deflected. On the other side this kind of configuration has also the disadvantage of being very sensitive to thermal noise, so it is necessary to balance appropriately these two opposite requirements as described in Sec. 1.3.4.

When the tip is approaching the sample, the interaction force between the tip and the substrate causes a deflection of the cantilever.

This force could have different origins: for example if the electrostatic is involved in imaging mode, the resulting force is due to Van der Waals interaction. In protein pulling experiments these interactions represent the resistance applied by the proteins to the unfolding process while in the imaging mode they reflect the interplay between the sample and the tip. This interaction results in a change of the position of the reflected laser beam on the photodiode, forming an *optical lever system*. The laser beam reflected by the tip hits a photodiode, which records the displacement (in Volts) of the beam from a reference point. In the imaging mode the laser displacement is converted to the distance between the tip and the sample, reflecting the height and the topography of the specimen while in force experiments the displacement is recorded as a voltage and subsequently converted to a force. By modelling the cantilever as a Hookean spring and by knowing its elastic constant, it is possible to convert the measured voltage to a force. Being capable to measure the deflection with nanometre precision, the force microscope is a force sensor in the picoNewton force scale. In practice, thermal noise for very soft cantilevers limits the force resolution to > 15pN. For the same reason, when the probe is retracted, if there are some interaction (adhesion, macromolecules attached between the tip and the substrate etc), the laser beam is deflected on the photodiode in the opposite direction. The intrinsic length resolution of the instrument, given by the precision of the piezoelectric actuator and by the sensitivity of the cantilever, is of the order of nanometres.

A piezoelectric actuator is able to move the cantilever in the 3 dimensional space with a resolution of one nanometre. By linking the spatial information to the force measured by the photodiode, several conclusions could be obtained, depending on the design of the selected experiment. When AFM is used in imaging mode, the probe scans the surface line after line. A proportional-derivative (PD) feedback instantaneously corrects the vertical position of the probe, ensuring constant applied force to the sample during the scanning.

At difference, when it is used for one-dimensional force spectroscopy the cantilever is moved only in the vertical direction, perpendicularly to the sample plane. In force spectroscopy the cantilever is moving up and down with a constant speed (from 30 nm/s to 3000 nm/s), and the position is recorded. There is also the possibility to maintain the probe at a constant bending (a constant deflection, or force, i.e force clamp mode, see Sec.1.3.2) by adjusting the tip position. In all these cases the goodness of the measure is strictly dependent on the feedback response time.

The AFM setup is particularly useful for the investigation of biological samples because a complicate preparation is not necessary and the measurements can be taken under near-physiological conditions (liquid buffer or controlled temperature).

In this Thesis the employed AFM is a commercial one, the Life Science Nanowizard® II from JPK Instruments, Berlin.

1.2.1 *Imaging*

The AFM was historically employed to characterize surfaces or for acquired images of biological specimens with a sub-nm resolution (in z-length) and with a lateral resolution of few nm. As explained above, when the tip of the cantilever is close enough to the specimen it is attracted or repelled from its surface, and this phenomena induces deflections of the cantilever. A *xy* raster scanning is employed to map the area of interest (usually from $100x100nm^2$ up to $100x100\mu m^2$, with a pixel range from 256 x 256 pixels to 8192 x 8192). The related variations in the voltage of the piezoelectric crystal controlling the vertical position of the cantilever are recorded and converted, in order to produce a 3D-image of the surface topography.

There are two classical strategies to acquire surfaces images: the contact mode and the tapping mode, even if nowadays several new modalities have been emerged (for example the Peak Force Tapping Mode, developed by Bruker®, see Appendix A.3).

Contact Mode

The first strategy to carry on AFM imaging experiments is the *contact mode*: the tip and sample are maintained in close contact during the scanning process, consequently they feel a repulsive Lennard-Jones potential. The major drawback of this imaging strategy is the presence of large lateral forces on the sample due to the dragging of the tip in contact with the specimen. These lateral forces often causes damages to the tip and to the sample, especially for soft biological specimens, while for crystalline surfaces these damages are generally not a big issue. In liquid mode the drag of the forces could limit the lateral resolution. The cantilever required have usually a low Resonance Frequency ($\sim 20kHz$) and a low spring constant ($\sim 0.1 - 1N/m$).

Tapping Mode

The second alternative imaging strategy is the *tapping mode*: the cantilever is oscillatingat a frequency near to the resonant one, in the range of hundreds kHz, so when it is approached to the surface of the sample the contact occurs only for a small fraction of its oscillation period. The interaction regime is still repulsive as the contact mode, but the very short times over which the contacts occurs dramatically reduce the lateral forces acting on the sample and on the tip. The cantilevers required for the non-contact imaging have usually an high Resonance Frequency ($\sim 300kHz$) and a high spring constant ($\sim 10 - 50N/m$).

Both the contact and the tapping mode can be applied both in liquid (with a low viscosity, like water and water-like medium) or in air. The advantages and the drawbacks of them are strictly dependent to the sample: usually the tapping mode is preferred for imaging biological samples due to its high sensitive and low perturbation of the sample, which prevents AFM-induced damages of the bio-macromolecule of interest. The liquid mode usually permits a lower resolution, due to the perturbation of the water molecule and a considerably lower resonance frequency of the cantilever, but has the significant advantage of study the sample in near-physiological conditions.

In this work, AFM imaging technique was widely employed, usually in tapping mode: during the analysis of the mechanical properties of the DNA analogue (Part 11.1, [2]), to observe the crowding effects on DNA induced by PEG and HNS (Appendix A.2, [9]), to characterize the aβ fibrils (Appendix A.1.1,[4] and A.1.2, [8]), to capture the morphology of collagen matrices (Appendix A.3, [3]) and to image cells (Appendix B.2) and fruits. Unless otherwise specified, I have personally acquired all the AFM images shown in this Thesis by means of the Nanowizard II instrument.

1.3 AFM- BASED SINGLE MOLECULE FORCE SPECTROSCOPY

Understanding the role of force in many biochemical process is still a challenging task, due to the necessary sensitivity in force needed for this kind of investigations. Indeed, a lot of different biomolecules are subjected to a force when performing their biological task, or their mechanical reaction to external force may want to be investigated. These experiments, which require an applied external force, are impossible with bulk techniques, especially if a specific knowledge of the direction and position of the force is needed. To overcome these problems, different apparatus have been developed. They all are single molecule based and the geometry of the experiment is carefully designed to have a perfect determination of the direction of the involved force.

In addition, the apparatus must have an high resolution both in length and force. In a typical single molecule technique for studying the stretching the of biopolymers, the biomolecule is tethered between its two ends. Then a force can be applied to one of the two sides and the reaction of the polymer is investigated. The AFM is an excellent technique in this kind of study for the reasons underlined above. As a matter of facts the probe is moved in the z-direction by the piezoelectric, first it is moved downwards until it gets in contact with the surface and then upwards, until there is no interaction between the probe and the surface. The resulting deflection of the cantilever (which is then converted into force) as a function of the tip-sample separation is called *force curve*.

The name *Force Spectroscopy* defines the presence of a mechanical force applied to the sample and the recording of the sample mechanical response to this force. Different Force vs Time (or Length) patterns could be recorded, in order to underline a different aspect of the sample's response to the external force.

1.3.1 *Force extension*

The most simple data that can be obtained by means of an AFM setup are called **Force Extension curve** (FX). In a typical FX curve (Fig 1.5), two different curves are seen: the approach curve (blue) and the retracted one (red).

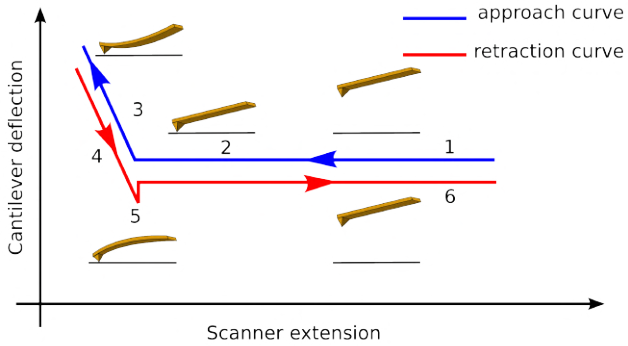

Figure 1.5.: A force-extension curve, showing the bending of the cantilever. The tip is moved downwards (1) until it gets in contact with the sample (2). A force is exerted on the substrate (3) for few seconds. Then the tip is moved away from the surface (4 - 5) until the starting position is reached again (6). (Fig. taken from [21])

If there is no interaction between the tip and the surface and the sample is sufficiently stiff, the retracting and the approaching curve are super imposable. Almost in all situations there is a dip in the retraction curve, whose depth could vary, (position #5 in the Fig.1.5): this represents the adhesion between the tip and the surface. It is more evident for soft and sticky samples, while for very clean, dry and stiff samples this feature could disappear.

The approaching speed of the probe toward the sample is defined by the operator: usually it is several hundreds of nm/s. Also the distance travelled by the tip depends of the kind of sample under study, from 500 nm to 10 µm (the maximum range spanned by the JPK Nanowizard is 15 µm).

When a single polyprotein is stretched (for a description of the employed polyproteins in force spectroscopy see Sec. 2.2), as the tip is withdrawn at constant speed, the extension of the molecule generates a restoring force that causes the bending of the cantilever.

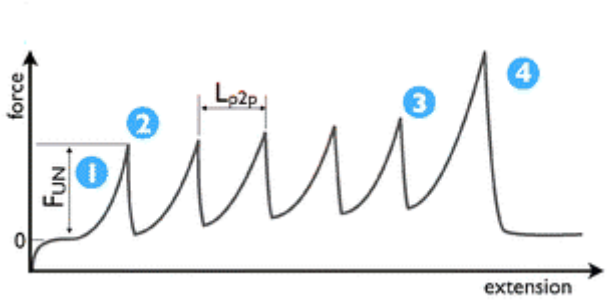

Figure 1.6.: A typical sawtooth pattern for the unfolding of a polyprotein in force extension mode (Fig. taken from [14])

During the polyprotein pulling, the restoring force acting on the cantilever increases until one domain unfolds: this unravelled domain suddenly adds its length to the total length of the protein and allows the force on the cantilever to fall. This phenomenon is detected as a peak in the force extension curve (Fig.1.6). After this, the protein resists to the elongation until the second domain in the chain unfolds. The retracted curve shows a variable number of peaks, each of them representing an unfolding event. This kind of curves is also called a *saw-tooth pattern curve* (Fig.1.6). When all the domains are unfolded, the polyprotein detaches from the tip and a detachment peak (whose height is usually bigger than other peaks) is observed (# 4 in Fig.1.6). This kind of curves are analysed by fitting each peak to the Worm Like Chain (WLC) function, and parameters as persistence length and contour length can be extracted (see Appendix D.3).

1.3.2 *Force clamp*

In Force Clamp (FC) spectroscopy , a single polyprotein tethered between the cantilever and the substrate is subjected to a constant force. In this way the extension length of the protein is measured directly as a function of time, giving rise to a **Force Clamp curve**. The main problem of this kind of experiments is that the AFM must have an excellent electronic feedback, in order to measure the force and to correct it instantaneously, by changing the vertical position of the cantilever. The typical required feedback response time is 2-6 ms. The more this time is short, the more the force is corrected quickly. In this way both the length and the force applied to the protein are precisely recorded as function of time.

For a polyprotein stretched at a fixed force, the extension vs time curve results in a well defined series of steps, each of them indicating one unfolding event. Indeed, the polyprotein length increases when a domain unfolds. As soon as one domain unfolds, the height of the cantilever must be corrected by the piezo to maintain the same force. This results in a step (Fig. 1.7) in the tip-sample separation vs

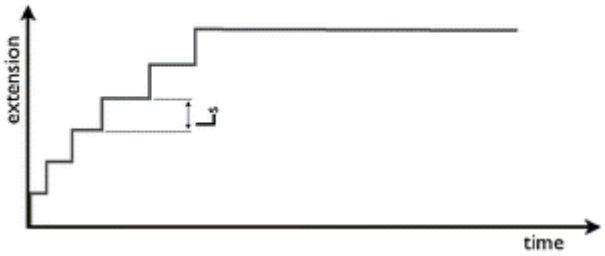

Figure 1.7.: A mechanical fingerprint for a polyprotein in a force clamp experiment. (Fig. taken from [14])

time curve. The final curve gives a *mechanical fingerprint* of the single polyprotein (Fig. 1.7).

From this kind of experiments, the length of a modulus can be found by measuring the height of a single step. It is also possible to find the rate of unfolding when the protein is subjected to different forces, by summing a high number of FC curves recorded in the same conditions and fitting it to a single exponential function.

The AFM-based force spectroscopy is used in this Thesis for unfolding a chimeric polyprotein ($I27_4 - \alpha Syn - I27_4$) in the Part ii.

1.3.3 *Elasticity*

The AFM-based nanoindentation has emerged as an useful tool to study elastic properties of biological samples, due to the versatility of soft cantilevers allowing local testing of fragile samples as cells (Appendix B.2.1, B.2.2 and B.2.3) tissue or substrates as collagen (Appendix B.1) [5, 6, 7, 26, 27, 28, 29, 30, 31, 32].

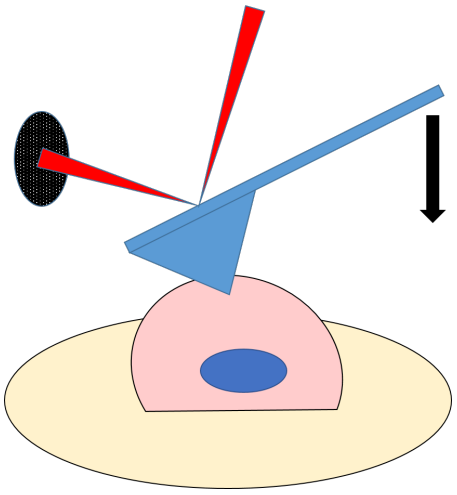

Figure 1.8.: A sketch of a AFM-based indentation on a cell.

The parameter of interest, the Young Modulus, which describes the elastic modulus of the sample, could be obtained by means of several different models, but most of them are based on the Hertz model and corrected to match the experimental conditions.

Hertz-Model

The Hertz model approximates the sample as an isotropic and linear elastic solid occupying an infinitely extending half space. Furthermore it is assumed that the indenter is not deformable and that there are no additional interactions between indenter and sample. If these conditions are met the Young's modulus (E) of the sample can be fitted or calculated using the Hertzian model. The force curves are plotted force (F) vs piezo displacement (d), and the Hertz model could be applied on the *extension* trace (blue trace in Fig.1.5). The hertz model assume an indentation smaller with respect to the thickness of the material, so it is valid only with an indentation depth up to 10%, which for a typical cell, whose diameter is approximately 5µm, is around 200-500 nm. In this way the substrate does not affect the measurement.

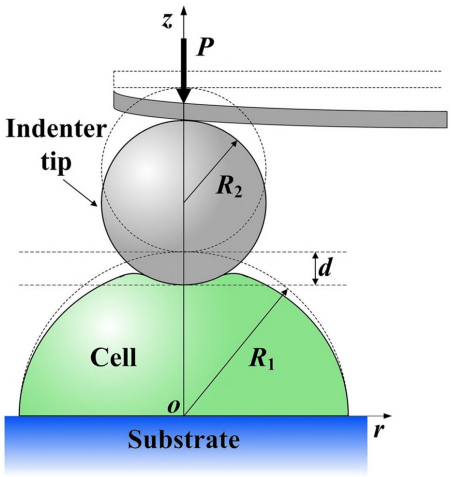

Figure 1.9.: The cell with radius R_1 is located on a rigid flat substrate and indented by a spherical indenter with radius R_2. (Fig. taken from [31])

The classical Hertz model considers two spherical bodies, while in a cell-indentation experiment the tip geometry has to be considered, since the contact surface is closely related to the geometry of the indenter [30].

The simpler model approximates the tip to a parabola:

$$F = \frac{4\sqrt{R_C}}{3} \frac{E}{1-\nu^2} d^{3/2} \tag{1.1}$$

where F is the force, R_C is the radius of tip curvature (R_2 in Fig 1.9), ν is the Poisson's ratio[2], E is the Young's Modulus and d is the piezo displacement.

1.3.4 *Cantilever Choice and Calibration for Force Spectroscopy*

Cantilevers for AFM measurements are typically made of silicon nitride using microelectromechanical systems, and they may be back side coated by gold in order to increase their reflectivity. This coating is essential for liquid measurements in order to increase the reflectivity of the laser and reduce water interferences. Different cantilevers could be mounted on a chip (Fig.1.10), and they could have one (rectangular shaped) or two arms (triangular or V-shaped), whose length is about $\sim 100 - 200\mu m$ (Fig.1.11). The shape of this arms is critical for the final calibration. In this Thesis both v-shaped and rectangular-shaped cantilevers were used.

For the protein unfolding experiments, the tip must be very sharp in order to maximize the probability of the attachment of a single molecule. For this reason, we used cantilever whose tip, mounted at the end of the cantilever arms, is a quadratic pyramid and has a curvature of tip radius of of $10 - 60nm$.

The choice of the cantilevers is critical to have good results, as they influence the signal-to-noise ratio, drift and feedback response time. As the laser beam waist of our setup is quite large (diameter $\sim 30 - 50\mu m$), it is necessary to have a cantilever with lateral dimension large enough to reflect most of the laser beam in order to avoid interference due to the reflection of the laser on the gold surface.

The smaller is the cantilever spring constant, the better is the noise-to-signal ratio (Fig.1.12). Indeed the deflection due to a given force is higher for a low spring constant cantilever than for high spring constant cantilever. Cantilever with small spring constant results in a better signal-to-noise ratio, because at the same force with small spring constant the cantilever is more deflected. On the other side, low spring constant cantilevers are more sensitive to the thermal noise. In addition, the resonance frequency of the cantilever limits the feedback response of the system. Therefore a good probe must be stable in a time of few milliseconds and must have a linear response in a quite large range of forces.

Each cantilever has a spring constant value declared from the producer and calculated from its geometry, but the spring constant value is very sensitive to imperfections that may occur during the production process.

Before choosing the more suitable cantilever for protein unfolding experiments, three different kind of cantilevers were tested: ORC-8,

2 the Poisson's ration depends on the material. For soft biological samples is ν=0.5, which describes incompressible materials like rubber.

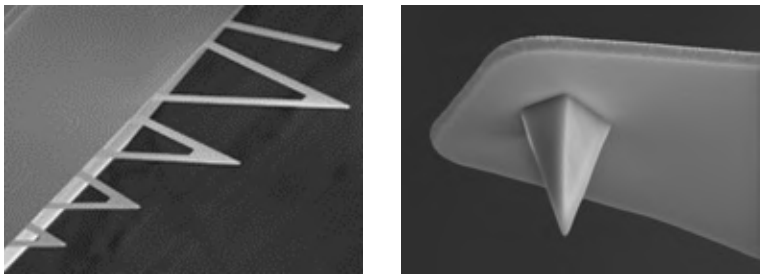

Figure 1.10.: Side B of MLCT probes and a zoom on the tip. The images are taken from Bruker official website.

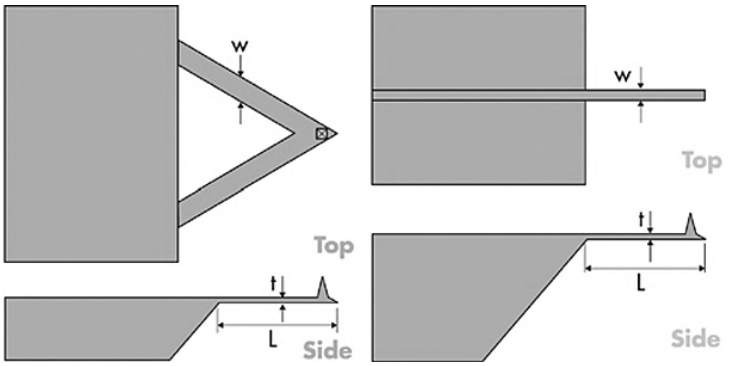

Figure 1.11.: Schematic examples of typical triangular and rectangular cantilever probes. The images are taken from Bruker official website.

DNP and MLCT, all of them are purchased from Bruker AFM probes (Table 1.2). The arms of these cantilevers have a length of $85-310\mu m$ and a width of $18-40\mu m$, while the tip is $\sim 2.5-8\mu m$ high.

The cantilever chosen for the α-synuclein unfolding is the MLCT-BIO # D (triangular shape, nominal resonant frequency of 7 kHz, nominal spring constant 0.03 N/m): it is soft enough to capture the signal coming from the protein unfolding with a good signal-to-noise ration in the liquid environment. It is golden coated and it has a nominal tip radius of 20 nm, large enough to assure the capturing of a protein spread on the surface but small enough to limit a multi molecule attachment.

For all the elasticity measurements (collagen matrices and cells reported in the Appendix B) the details of the cantilever chosen were described individually.

Cantilevers

MLCT

Name	Shape	Resonant freq kHz	Spring constant N/m
B	Rectangular	15	0.02
C	Triangular	7	0.01
D	Triangular	15	0.03
E	Triangular	38	0.1

ORC-8

Name	Shape	Resonant freq kHz	Spring constant N/m
LN	Rectangular	18	0.05
SN	Rectangular	68	0.38
LW	Rectangular	19	0.1
SW	Rectangular	71	0.73

DNP-S

Name	Shape	Resonant freq kHz	Spring constant N/m
LN	Triangular	18	0.06
SN	Triangular	56	0.24
LW	Triangular	23	0.12
SW	Triangular	65	0.35

Table 1.2.: Table of the cantilevers used in the experiments. The acronyms L (long), S (short), N (narrow) and W (wide) are here used for distinguishing the different cantilevers mounted on the same chip. The values reported for the resonant frequencies and for spring constants are the ones declared by the producer. The discrepancy between these values and the ones calculated by thermal calibration is about $\pm 30\%$ for soft cantilevers and less than $\pm 15\%$ for the cantilevers with high spring constant (ORC-8 SW and DNP-S SW)

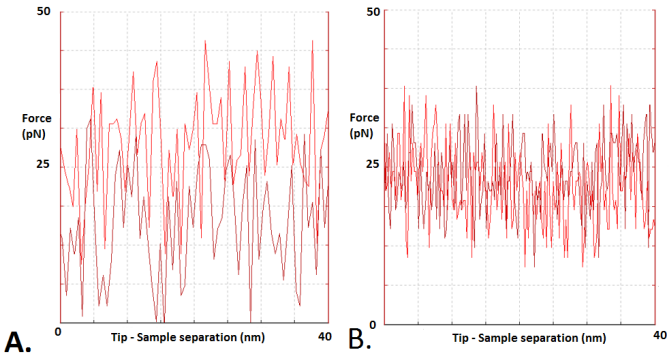

Figure 1.12.: Zooms of different force vs. distance curves acquired in buffer in the non interaction regions. The cantilever A. has a $k = 0.05\text{N/m}$ and a thermal noise of $n_{thermal} \sim 25\text{pN}$, while the cantilever B. is five times stiffer ($k = 0.25\text{N/m}$) and has a thermal noise of $n_{thermal} \sim 10\text{pN}$)

Cantilever calibration

The bending of the cantilever, and consequently the displacement of the laser beam on the photodiode, is directly proportional to the tip-sample interaction force. The change in voltage measured by the photodiode can be converted into a force by a suitable calibration procedure. This step is critical to obtain reliable results.

First, the measured change in voltage (Volt) needs to be converted in the deflection of the cantilever (nm). This conversion factor, called *sensitivity*, depends on the cantilever, on the optical path of the AFM detection laser and on how the cantilever is mounted on the glass holder. In order to obtain this value, a force calibration extension curve is recorded. The cantilever tip is brought close to a hard surface and pushed, and deflection-extension data are obtained (Fig.1.13). When the probe is in contact with the surface and the piezoelectric is approaching with a constant speed, the displacement of the laser beam on the photodiode is directly proportional to the physical displacement of the piezoelectric. The multiplicative inverse of the slope of the force-curve in the contact region shown in Fig.1.13 yields a correlation between the change in voltage measured by the photodiode and the bending distance measured from the movement of the piezoelectric (sensitivity $= \Delta z/\Delta V$). The sensitivity is expressed in units of nmV^{-1}.

For small bending angles, the cantilever behaves like an ideal Hookean spring with a spring constant (k) that is determined by its material and dimensions. This allows using the Hooke's law:

$$F = k \cdot \Delta z \tag{1.2}$$

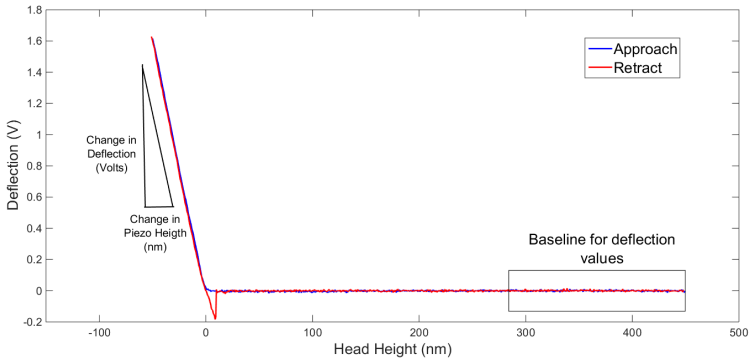

Figure 1.13.: Measured Approach (blue) and Retract (red) deflection versus distance of the height data in air for sensitivity calibration. In this case the sensitivity is 41nmV^{-1}. Calibration data acquired with cantilever MLCT-C.

By using the Eq. 1.2 the deflection of the cantilever (in nm) is then converted in force (pN) (Outline 1.3).

$$\text{Force} = \text{verticaldisplacement} \cdot k = \text{deflection} \cdot \text{sensitivity} \cdot k \quad (1.3)$$

In order to measure the spring constant of the cantilever k, different methods are reported in literature [33], but the most common is the **Thermal Fluctuations method**, which is based on the equipartition theorem. Even if this method is less precise than others, such as the method of calibration with a reference cantilever, it is widely used because it is the easiest and faster one. In addition, this method is the most indicated for the soft cantilevers calibration.

The end of the cantilever is constantly fluctuating because of the thermal vibrations from the environment. These fluctuations can be modelled as a kind of diffusion restricted by the restoring force due to the spring constant.

As a consequence of the equipartition theorem, which states that the kinetic energy of each degree of freedom (that in our case is the vibrational mode) equals half the thermal energy $k_B T (= 4.11\text{pN} \cdot \text{nm})$, the constant spring is obtained. So, in the final calibration step, the thermal fluctuations of the cantilevers far from the surface are measured as a function of the frequency (cantilever thermal noise spectrum) (Fig.1.14). The vertical deflection of the cantilever (z_c in Eq. 1.4) is measured along some time. This process may be represented by an histogram of vertical deflection values. Nevertheless, it is easier to look at the frequency dependence of the fluctuations, and this procedure intrinsically excludes the low frequencies or specific noise sources. On average, the greatest amplitude will be found around the cantilever resonance frequency.

The amplitude of these fluctuations at a given temperature (the room one) depends only on the spring constant of the cantilever. The

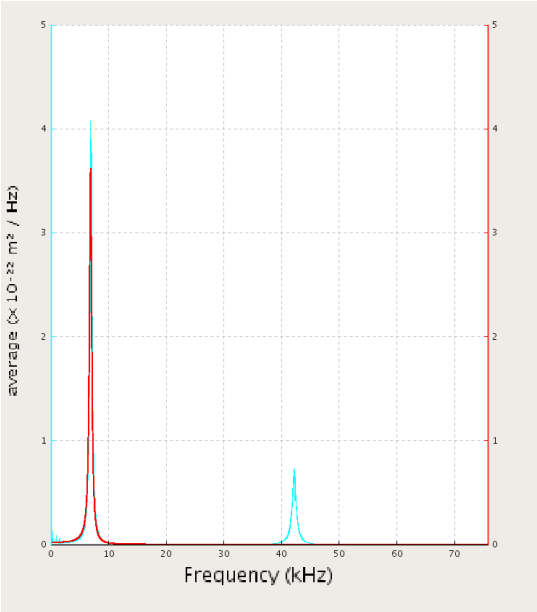

Figure 1.14.: Measured resonance spectrum of a soft cantilever MLCT-C in air. Two peaks are shown: the first one ($f_{max} \sim 6.8$kHz) is the first vibrational mode, while the second ($f_{max} \sim 44$kHz) represent the second mode. The two spring constants are $k_{1st-mode} = 11.52$mN/m and $k_{2nd-mode} = 10.06$mN/m. The red line is the Lorentzian fit of the first vibrational mode.

thermal resonance curve can be therefore fitted by a Lorentz function, and the area under the curve is used as a measure of the resonance energy in order to obtain the value of the square mean cantilever vertical deflection in z $\langle z_c^2 \rangle$.

If the cantilever is considered as a simple harmonic oscillator and the Brownian motion of the fundamental oscillation mode of the cantilever is evaluated, the elastic constant is given by the simple formula:

$$k = \frac{k_B T}{\langle z_c^2 \rangle} \tag{1.4}$$

where k_B is the Boltzmann constant and T is the temperature in Kelvin.

Actually, considering the cantilever as an ideal spring, the resulting spring constant has a typical error of $\pm 20\%$ with respect to the real value and this error is not negligible in the range of forces that we explored. Indeed, the real cantilever is not completely harmonic as assumed by the theory. Actually there are various correction factors that must be used to get a more accurate value from the fit. There are complex theories, that have derived empirical equations to correct the signal [34]. The disadvantages are that such theories require a precise knowledge of the environmental conditions (Temperature, refraction index of the mean..) and of the geometry of cantilever (including the purity of the material). A compromise was found by introduc-

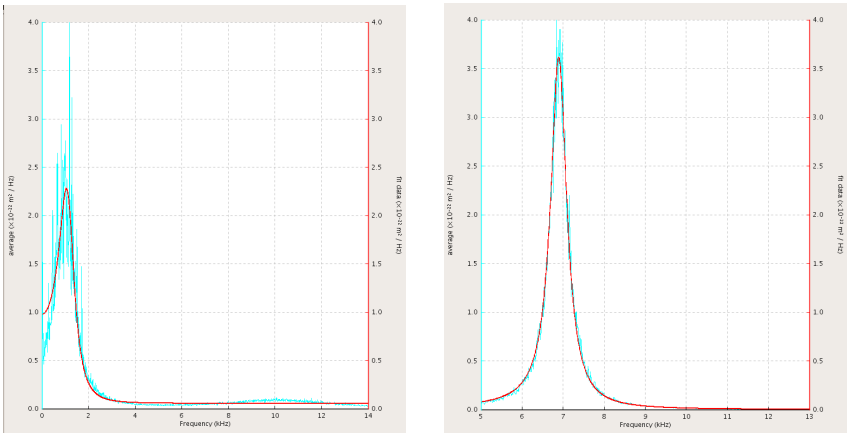

Figure 1.15.: Measured spectra in liquid (left) and air (right) for the same MLCT - C cantilever. The spring constants are $k_{liquid} = 10.09$ mN/m and $k_{air} = 11.52$ mN/m. Red lines are Lorentzian fits.

ing *correction factors* dependent to the shape of the cantilever. These *correction factors* are found by deriving an analytical estimation of the spring constant by means of a finite element analysis.

The correction factors for the first vibration mode are $f = 0.764$ for triangular cantilevers [34] and $f = 0.817$ for rectangular ones [35]. These values are valid only if the laser is positioned at the very end of the cantilever: if the beam is moved toward the chip the predictions are no more reliable.

Fluid and Air calibration

In theory, the calibration procedure described above is valid both in liquid and air. However, when using soft cantilever, such as the ones described in the Table 1.2, the determination of the spring constant can be problematic. Indeed, very soft cantilevers, such as MLCT-BIO C or D ($k \simeq 0.01 - 0.03$N/m), are strongly affected by the viscosity of the liquid and their spring constants may be underestimated. For these cantilevers the difference in the spring constants obtained by the thermal fluctuations method in air is about $15 - 20\%$ higher than the same calibration performed in liquid.

For an accurate calibration, both the two steps of the procedure described above are performed in air. In this case the resonance peak is more easily detectable, because the thermal noise is less visible with respect that in fluid and it has a frequency high enough to be clearly distinguishable from the low-frequency noise (Fig.1.15).

The sensitivity must be calculated anew at the beginning of the experiment (in liquid) while the spring constant is the same of the one determined in air.

The Magnetic Tweezers are the most sensitive technique among all the single molecule low-force spectroscopy set up. The first magnetic set up to manipulate DNAs was developed by Amblard et al. [36], while the first modern Magnetic Tweezers (MT) setup was developed by Strick and co-workers [37].

The typical MT experiment consists on tethering the molecule of interest (historically DNA, but in the last five years the MT-based force spectroscopy of polyproteins has emerged [38]) by means of a paramagnetic bead, which can be manipulated using a pair of permanent magnets placed above the specimen. The commercially available paramagnetic beads have a 1-10 µm diameter, and they produce a characteristic diffraction pattern: optical images are acquired by an inverted microscope, permitting to track the position of the bead. The extension of the molecule, and the force the bead is subjected to, can be determined from the bead's coordinates, approximating the DNA-bead to an inverted pendulum.

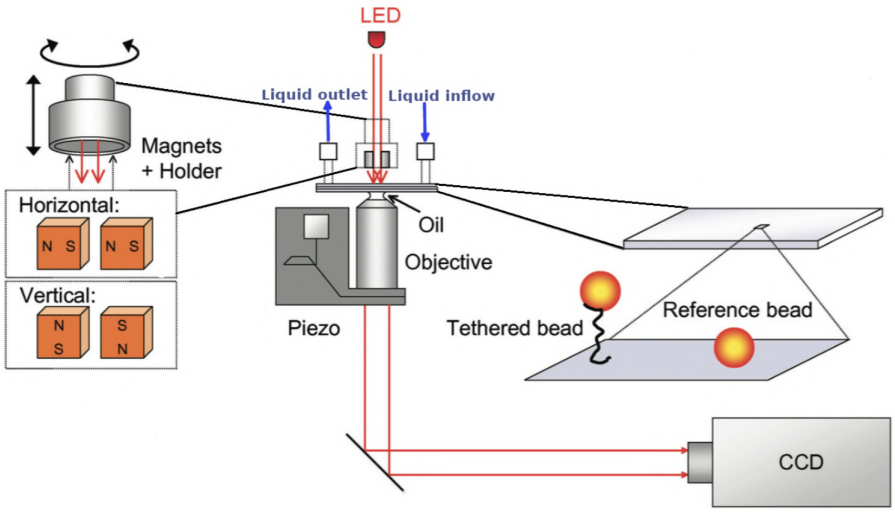

Figure 1.16.: Scheme of a typical magnetic tweezers setup. The two possible configurations of the magnets are reported, the horizontal geometry is that one used in this work. Figure adapted from [49].

Among single molecule techniques the MT have several advantages: first of all it is the most straightforward and cheapest set up to build, it has a simple and robust experimental configuration and it is intrinsically in force clamp configuration, without the necessity to have a feedback control on the forces. Moreover, the resolution of the force applied to the molecule is very precise, in the range from 10^{-3} to 10^{2} pN, depending on the iron content of the beads employed. The MT set up can be easily applied for multiplexing measurements, permitting the simultaneous measurements of several

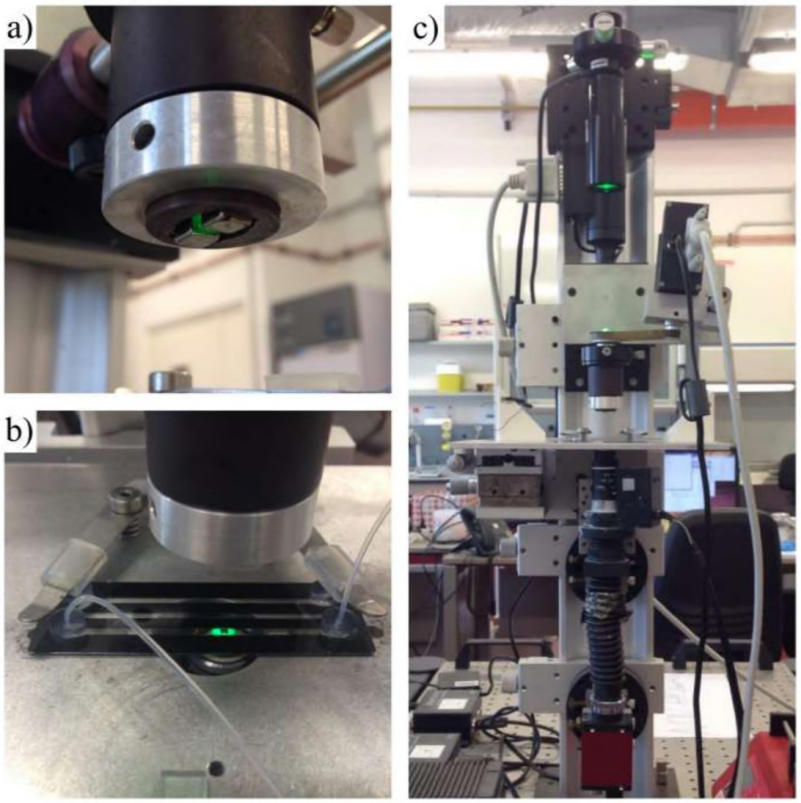

Figure 1.17.: Pictures of the MT setup in the Mantegazza Lab (Chapter 11.1).

tens of DNA molecule with the same single molecule resolution, overcoming the limitation of slow and time consuming single molecule experiments. Moreover, this set up can easily improved with fluorescent microscopy (as TIRF or confocal microscopy).

Finally MT is capable of inducing a torsion to DNA molecules by rotating the pair of permanent magnets and finely monitoring supercoiling dependent processes.

The MT-based experiments reported in this Thesis have been acquired by means of two different home-build setup, one in Francesco Mantegazza Lab (Chapter 11.1, Fig.1.17, details in [2, 40, 41, 42] and in the Moreno-Herrero Lab (Part iv and v, details in [43, 44, 45, 46]).

1.4.1 *MT set up*

A MT standard set up can basically work through two operation principles: *Pulling force experiment* in which magnets have a fixed position and they are moved up and down by means of a piezo. In this way the force applied to the macromolecule varies as a function of the magnets position. In the second configuration, the so-called *Twisting experiment*, the two magnets are held at constant height, i.e. keeping the force constant, while they are rotated to apply a torque to the

torsionally constrained DNA molecules. All the MT-data reported in this Thesis are obtained in the pulling force mode.

The set up of the MT at the Mantegazza Lab (structural simile to the one employed at the Moreno-Herrero Lab, detailed described in Sec. 13.4) consists of an optical inverted microscope equipped with an oil-immersion objective (NIKON 100x, $NA = 1.25$) mounted on a piezoelectric focusing system (PIFoc, Physik Instrumente, Italy). The objective is coupled with a 15cm focal-length lens leading to a real 75x magnification. The light source is a green light-emitting diode (LED). The charge-coupled device (CCD) camera (Marlin, Allied Vision Technologies) collects images at a frame rate of $50 - 100$Hz and it is connected to a computer to analyse images in real time to obtain the magnet position (then converted into the force) vs bead position along z-axes.

The magnetic field is generated by a micro-movement apparatus of two permanent magnets separated by a small gap (i.e $1 - 2mm$). Two motors control the position of the magnets along the optical axis (z-direction). The magnets are usually two cubes of 3 of neodymium iron boron ($Nd_2Fe_{14}B$). The magnetic field decreases approximately exponentially moving away from the magnets, therefore generating a gradient field along the optical axis [47].

The flow chamber consist in a gasket of parafilm sandwiched into two coverslips, one of them is drilled so that the gasket creates a channel in which introduce the sample (DNA) and then other elements (i.e. proteins, crowding agent, a measurement buffer).

The DNA must be tethered between the flow cell lower coverslip and the paramagnetic bead. To obtain this result, the most common option is to obtain a double functionalization: digoxigenin-antidigoxigenin and biotin(strept)avidin interaction. Paramagnetic beads labelled with streptavidin are commercially available and (Dynabeads, MyOne streptavidin) and the lower glass coverslip of the camber must be properly functionalized with anti-digoxigenin [2, 44]. In the most of cases, a proper passivization of the flow chamber can be required, by using polystyrene or BSA (Bovine-Serum-Albumine) protein. The DNA molecule is then labelled with biotins in one end and with digoxigenin molecules on the other one using either PCR fabricated DNA handles or labelled oligonucleotides [42].

1.4.2 *Optical Calibration*

By employing a MT setup, it is possible to obtain the extension of DNA molecules in real time, using an optical calibration method based on the circular diffraction pattern generated by the magnetic beads tethered to DNA [48]: basically different diffraction images are acquired by means on the CCD camera by shifting the objective along the z-axial direction, with small and precisely determined steps. This

process creates a series of diffraction rings of the magnetic beads, generating a calibration profile by correlating the diffraction pattern to the distance of the bead from the objective focal plane (Fig.1.18). The difference between two measured distances gives the vertical extension of a DNA molecule, with an accuracy dependent of less than 10nm. This calibration procedure enables to determine the axial position with an accuracy up to ~ 10nm.

In a MT experiment, at least two beads are used: one stacked to the surface serving as a reference and the other one tethered to the DNA molecule, and the optical calibration is performed for both of them (Fig.1.19, Right). The advantage of a differential measurement between these two optical calibration profiles is the strong minimization of thermal drift effects or instrumental noise.

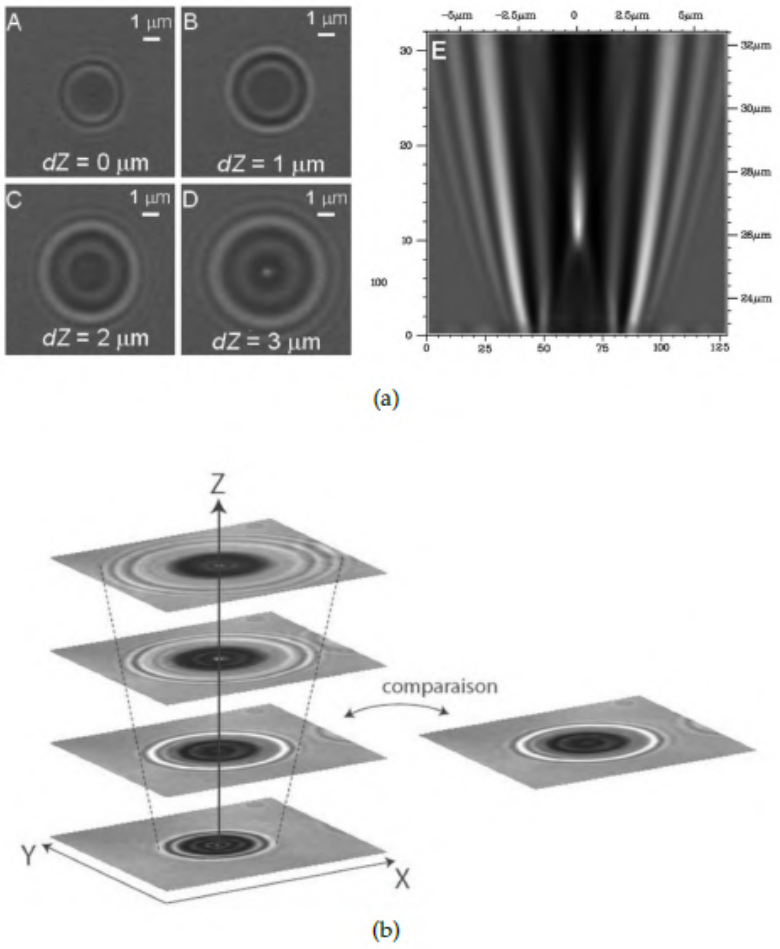

Figure 1.18.: (a) Left: diffraction rings formed by a 3-m magnetic bead at four different positions along the axial direction. Right: Calibration profile related to the intensity of the diffraction rings at several positions. (b) Schematic representation of the stack of images recorded by the camera. Figures adapted from [49].

The x-y fluctuation of the DNA beads fluctuations of the beads are necessary also to calculate the applied force, using a cross-correlation

analysis of the calibration intensity profiles in the horizontal plane. The tracking on the x-y plane can be carried out with the same accuracy of the vertical one, with a resolution of a few nanometres.

1.4.3 *Force Calibration*

The magnets create a magnetic field that is applied to the paramagnetic beads, thus generating the following force

$$F = \frac{1}{2}\nabla(\mathbf{m} \cdot \mathbf{B}) \tag{1.5}$$

where $\mathbf{m}$ is the magnetic moment of the bead induced by the external magnetic field $\mathbf{B}$. The variations of the magnetic field are in the range of millimetres, consequently it can be considered constant for the micron-sized magnetic beads.

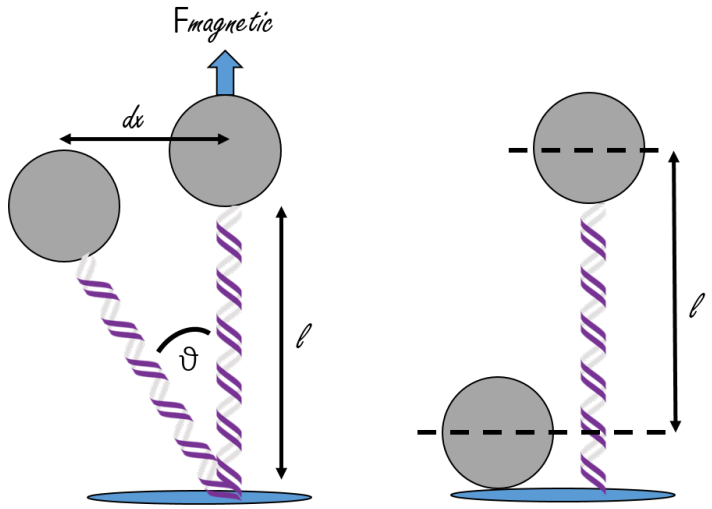

Figure 1.19.: Left: The DNA-tethered bead is considered to behave as an inverted pendulum under Brownian fluctuations. Right: Extension (l) is determined from the center of a tethered bead to the center of a reference bead located on the surface.

By using the Equipartition Theorem, the DNA-thethered-bead could be approximated as an inverted pendulum system under Brownian Fluctuations (Fig.1.19):

$$\frac{1}{2}k_x\langle dx^2\rangle = \frac{1}{2}k_B T \tag{1.6}$$

By using the approximation at small angle ($sin\vartheta = \vartheta = dx/l$), the final force is $F = F_{magnet}\,dx/l$ leading to the final expression of

$$F_{magnet} = \frac{k_B T l}{\langle dx^2\rangle} \tag{1.7}$$

Consequently, the force which the molecule is subjected to could be obtained just by measuring the DNA extension (l) and the transversal (i.e. in the x-y plane) fluctuation (dx).

The standard procedure consist in recording the x-y-z track of a DNA bead at different magnet position and then the force is plotted against the magnet position, which is fitted to an exponential function to obtain an empirical calibration curve.

Because the DNA filaments can be approximated as polymers tethered by a force, their behaviour could be describe by a polymer-physic model, the Worm Like Chain (WLC) (See Appendix D). By using this model, parameters describing the DNA could be extracted: the persistence length (L_P), which is related to the stiffness of the DNA and it is usually around 45-50 nm under standard conditions, and the contour length (L_C), which is the length of the studied molecule and is $L_C = 0.34nm/bp$ [50].

In the last 10 years new single molecule approaches for exerciting forces to macromolecules has emerged in the scientific community [51] to provide key information on the dynamic kinetics, and heterogeneity of proteins and molecular motors involved in the organization of DNA [52]. Among these, one of the most widespread is the *Flow Stretching* combined with *Fluorescence Microscopy* applied to DNA molecules [51, 52, 53, 54, 55].

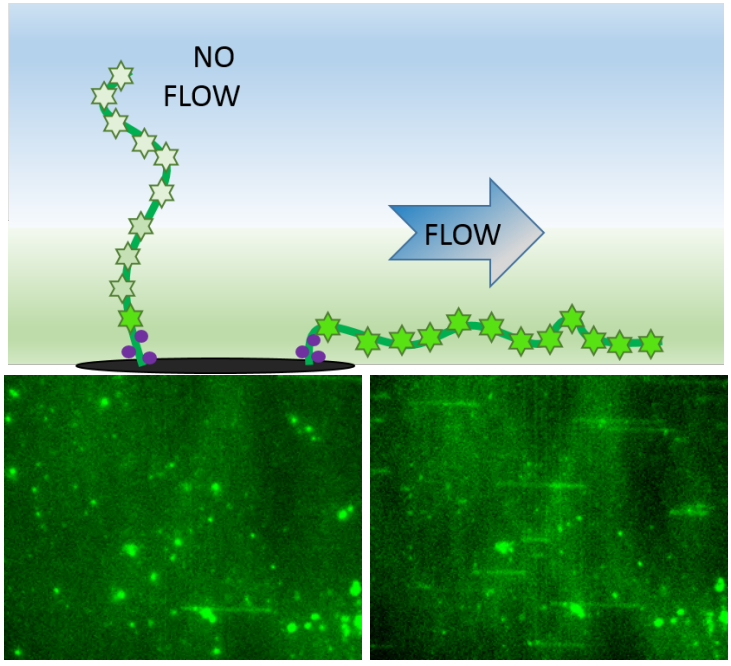

Figure 1.20.: Top: Cartoon of a typical flow Stretching (F-S) coupled with fluorescence experiment on DNA molecule.
Bottom: Real images of fluorescent DNA molecules labelled with SitOx green intercalant without flow (Left) or with an applied flow rate Q = 50 µl/min (Right).

This approach is particularly suitable to directly visualize protein - DNA interaction, both manipulating the samples and imaging the DNA-protein complex through direct labelling with fluorescent compounds (fluorescent labelled DNA or protein, fluorescent intercalating agents or Quantum Dots-conjugate to proteins or nucleic acids). Several instruments have been developed to couple the single molecule manipulation with TIRF (Total Internal Reflection Fluorescence) microscopy: optical tweezers [56], Lateral MT [44] or flow stretching [51].

Particularly, the latter approach stretches the DNA molecules by taking advantages of an external flow, which could be applied by a syringe pump mechanism. This approach is very easy to be implemented in a standard MT-TIRF apparatus without substantial modification.

In brief, the external flow applied to DNA stretches the molecule along the flow cell surface, in the visualization plane, permitting its observation by fluorescence. This result could be achieved by using a DNA molecules with a single functionalisation, resulting in a fully stretched DNA, or by using a molecule functionalised at both the ends, to monitor, for example, the DNA supercoiling [53, 57] or protein DNA loop extrusion [54, 58].

In this Thesis the Flow-Stretching - TIRF measurements (Part v) has been implemented in the Moreno-Herrero Lab in a setup described in [44] (Fig.1.21).

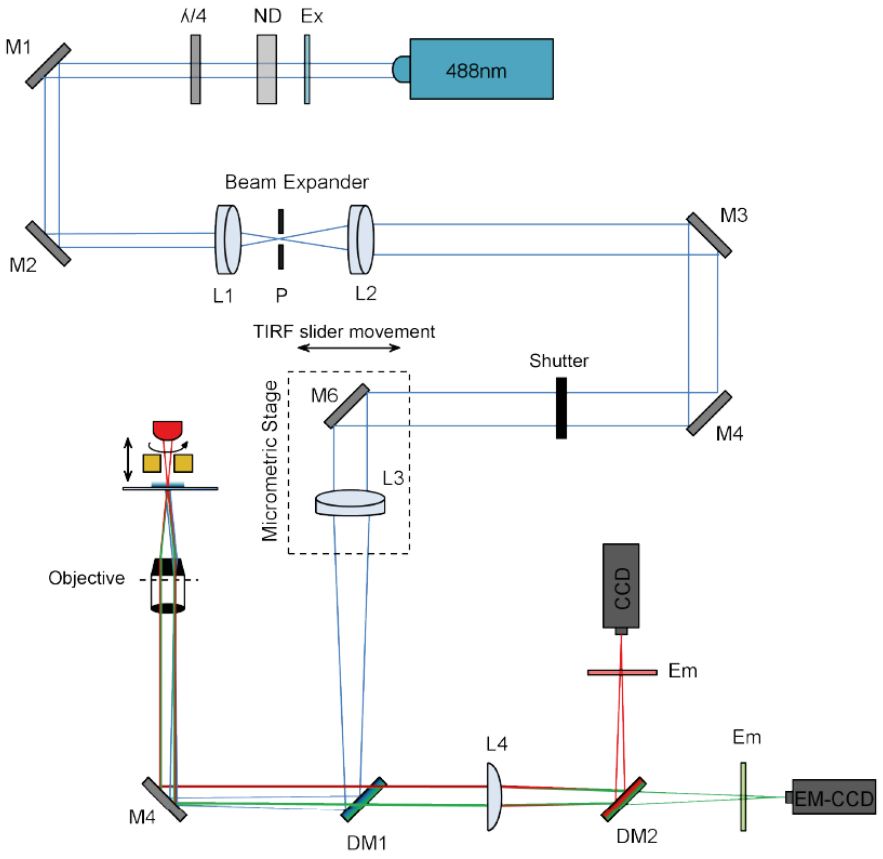

Figure 1.21.: A 488 nm (blue) linearly polarized light (Vortran Stradus) is filltered, attenuated, circularly polarized to avoid preferential excitation of fluorophores and expanded about 10 times for homogeneity of the beam. A mirror (M6) placed on a micrometric stage allows switching between TIRF and EPI-illumination, by translation, while the lens L3 focuses the beam on the BFP of the objective (Olympus UAPON TIRF 100x). Light from the tweezers LED (red) and emitted fluorescence (green) go back to the objective, passes through dichroic mirror DM1 and is focused on an Andor Ixon Ultra 897 EM-CCD camera (fluorescence emission) and CCD camera (LED light) by L4. Dichroic mirror DM2 allows separating both beams. Figure taken from [62].

1.5.1 *Total Internal Reflection Fluorescence Microscopy*

Several methods have been applied to fluorescence microscopy to restrict the detection of fluorophores to a thin region of the specimen, as super resolution techniques (STED) to confocal microscopy to non-linear excitation microscopy. Limiting the background from outside improves the resolution of the image and dramatically improve the signal-to-noise ratio.

Total Internal Reflection Fluorescence Microscopy (TIRFM) take advantage of an induced evanescent wave in a limited specimen region immediately adjacent to the interface between two media having different refractive indices. In practice, the most commonly utilized interface in the application of TIRFM is the contact area between a specimen and a glass coverslip, which in the DNA-Flow Stretching experiments is the lower surface of the flow cell.

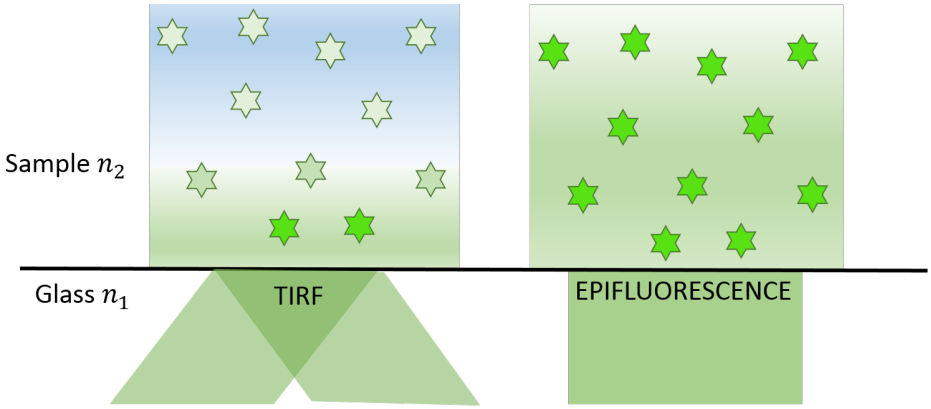

Figure 1.22.: Comparison between TIRF and Epifluorescence. In Epifluorescence the light illuminates the whole sample (along the optical axis), while in TIRF the evanescent wave only reaches a depth of $\sim 100 - 200nm$ end excites only fluorophores close to the glass surface.

Physical Basis of TIRF

The physical phenomenon which describes the total internal reflection (TIR) is at the base of multiple applications, as diamond cutters to optical fibers.

The refraction of light as it encounters the interface between two media having different refractive indices (n_1 and n_2, for example glass and air or glass and water) results in confinement of the fluorescent wave to the higher-index medium.

A collimated light beam (for example, a laser beam) propagating through one medium and reaching such an interface is either refracted as it enters the second medium, or reflected at the interface, depending upon the incident angle and the difference in n_1 and n_2.

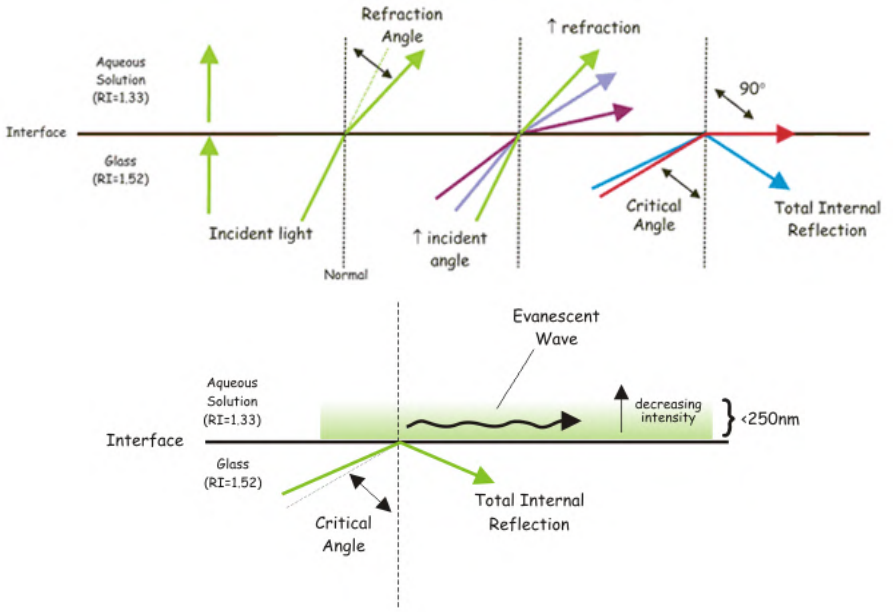

Figure 1.23.: (Top) Below the critical angle ϑ_c a fraction the incident light is reflected and a fraction transmitted. After reaching the critical angle, all the light is reflected entering TIR, and generating an evanescent wave [63]. (Bottom) The evanescence wave used in TIRF microscopy. Figures taken from Robarts Research Institute website.

Total internal reflection is only possible in samples in which the propagating light encounters a boundary to a medium of lower refractive index, according to Fresnel's laws. In dielectric media the angles of incidence and refraction of an electromagnetic beam is governed by Snell's law.

In standard conditions, part of the incident light is reflected and part is refracted, but in TIR condition the angle is adjusted to eliminate the transmitted component of the light ($\vartheta_2 = 90^o$ and $\sin\vartheta_2 = 1$) and the critical angle (ϑ_C) can be directly calculated from Snell's equation:

$$\vartheta_C = \sin^{-1}\frac{n_2}{n_1} \tag{1.8}$$

For all the incident angles satisfying the condition $\vartheta > \vartheta_C$, the total internal reflection is produced (Fig.1.23, Right). In the setup used in Chapter v the two media used are silica glass (n_1=1.52) and water ($n_2 = 1.33$), with a consequent critical angle $\vartheta_C = 61.7^o$.

1.5.2 *Taylor-Aris Theory*

In this Thesis the FS-TIRF setup has been used with a two computer-controlled syringe pumps, in order to achieve a fast buffer exchange under the condition of laminar flows and minimal transversal diffusion.

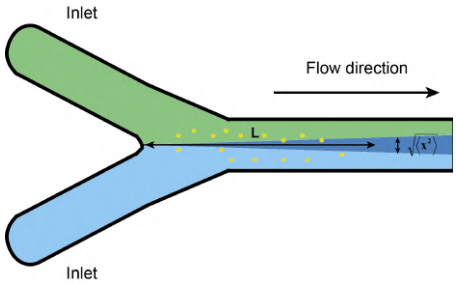

Figure 1.24.: Particles diffusing in flow cell (yellow), the further from the inlets, the bigger the transverse diffusion (Figure adapted from [62]).

To characterize the quality of the boundary exchange, the buffer solution is periodically switched with a 5mM Fluorescein aqueous solution and the images are recorded in the proximity of the surface (less than 150nm from the lower cell coverslip). As reported by Niman *et al* [59], the fall time t_s is defined as the time to go from 90% intensity to 10% intensity. The measured values can be compared to the Taylor-Aris model [60, 61]. This theory is valid where this condition is satisfied:

$$\frac{y}{v_0} >> \frac{a^2}{288D} \tag{1.9}$$

with y is the distance from the entrance of the interface, v_0 is the velocity of the fluid and a the height of the channel. The Taylor-Aris is valid at distances $y >> y_0$ from the two-inlet junction (in our setup $y \sim 1-2mm$). By using the effective diffusion coefficient k

$$k = D + \frac{\kappa(av)^2}{D} \tag{1.10}$$

If a rectangular channel is considered, $\kappa = 120$, the final expression of the fluorophore concentration as function of time:

$$c(t) = c_0 \frac{1}{2}\left(1 + Erf\left[\frac{y - vt}{2\sqrt{tk}}\right]\right) \tag{1.11}$$

where c_0 is the concentration of fluorophores.

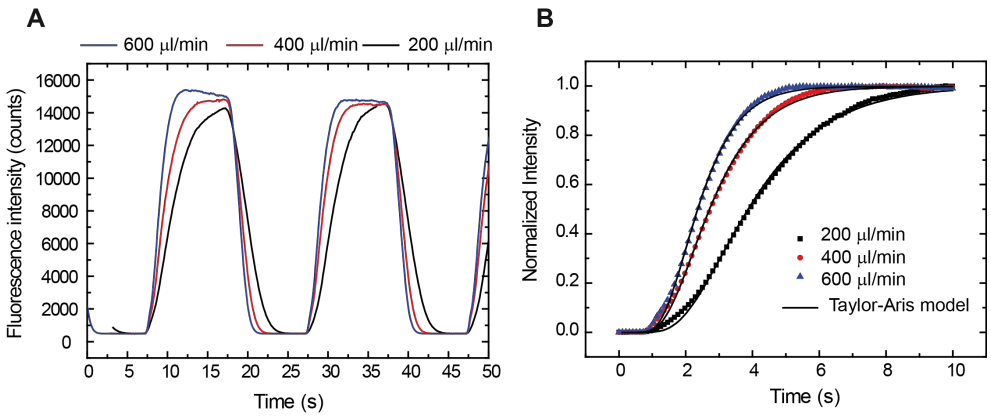

Figure 1.25.: A) Cycles of buffer/fluorophore exchange at three different flow rates showing that exchange rates range from 3-6.5 s (90% total intensity) in a 2 mm - 2 layer parafilm flow cell. B) Data were well fitted to the Taylor-Aris model consistent with a change of boundary. Figure taken from [43].

2

PROTEINS

2.1 BIOLOGY OF PROTEINS

Proteins are the most abundant biological macromolecules, occurring in all cells and also in all sub-cell structures. They also occur in great variety: in the same single cell there are thousands of different kinds of proteins, ranging in size from relatively small peptides to huge polymers. Surprisingly, all proteins are composed by relatively simple monomeric subunits, called amino acids, that provide the key to the structure of proteins which have a very big variety of functions. Indeed, proteins perform a vast array of functions within all the living organisms, including DNA replication, catalysing metabolic reactions, responding to external stimuli, providing structure to cells and transporting molecules. All proteins, whether from the most ancient lines of bacteria or from the most complex forms of life, are constructed from the same ubiquitous set of 20 amino acids, covalently linked in characteristic linear sequences [64].

2.1.1 *Amino acids*

Proteins are polymers of amino acids, in which each amino acid residue is joined to this neighbor by a typical covalent bond, the so called *peptide bond*[3]. All 20 of the common amino acids are β amino

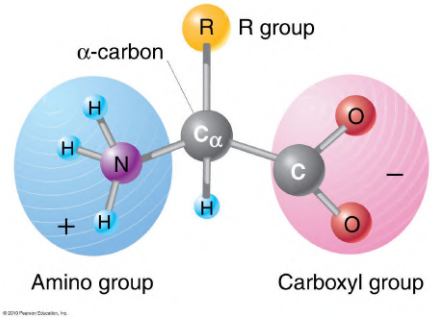

Figure 2.1.: The amino acid structure. (Fig. ®2010 Pearson Education, Inc)

3 A *peptide bond* is a chemical bond formed between two molecules when the carboxyl group of one molecule reacts with the amino group of the other molecule, releasing a molecule of water, resulting in an CO−NH bond

acids: they have a carboxyl group and an amino group bonded to the same carbon atom (the α carbon). They differ from each other in their side chains, or R groups, which vary in structure, size, electric charge and which influence the solubility of the amino acids (Fig.2.1). Amino acids are usually classified by the properties of their side chain into four different groups: (1) *non-polar amino acids*, also called *hydrophobic*, if the R group is either alkyl or aromatic, (2) *polar amino acids* or *hydrophilic* if the side chains contain different polar groups like amines, alcohols or acids. They can be further divided into (2a) *acidic*, (2b) *basic* and (2c) *neutral* [65].

2.1.2 *The protein structure*

To describe a complex macromolecule, such as a protein, it is necessary to define different structural levels, arranged in a sort of conceptual hierarchy. Four levels of protein structures are commonly defined (Fig.2.2): a description of all covalent bonds (mainly peptide bonds and disulfide bonds) linking amino acid residues in a polypeptide chain is its *primary structure* so, it represents the sequence of amino acid residues. *Secondary structure* refers to particularly stable arrangements of amino acid residues giving rise to recurring structural patterns, such as α -helix or β - sheet. *Tertiary structure* describes all aspects of the three-dimensional folding of the polypeptide. When a protein is composed by two or more polypeptide subunits, their arrangement in space is referred to as *quaternary structure*.[64]

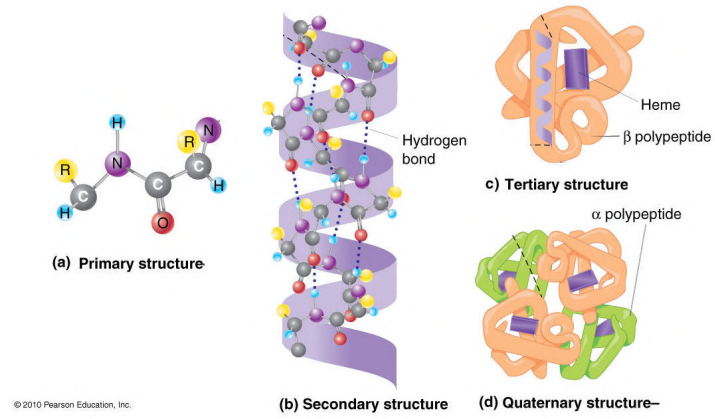

Figure 2.2.: The four levels which describe the protein structure. (Fig. ®2010 Pearson Education, Inc)

2.1.3 *The three dimensional structure of proteins*

The covalent backbone of each protein is formed by hundreds of individual bonds. If we consider this protein as a chain of amino acids without any constriction, so able to freely rotate in the 3-dimensional

space, the protein can assume an unlimited number of conformations. However, by the late 1920s, several proteins have been crystallized: the molecular structure was identical for all the proteins of the same specie. Proteins have a 3-dimensional structure, depending to their functions in the cell, which is directly correlate to their specific amino-acids sequence.

The five conclusions that emerge from this kind of study are (according to [64]):

- The three-dimensional structure of a protein is determined by its amino acids sequence.
- The function of a protein depends on its structure.
- An isolated protein usually exists in one or a small number of stable structural forms.
- The most important forces stabilizing the specific structures maintained by a given protein are non-covalent interactions.
- Among the huge number of unique protein structures, it is possible to recognize some common structural patterns that help in the understanding of protein architecture.

These themes should not be taken to imply that proteins have static, unchanging three-dimensional structures: in fact protein function often entails an interconversion between two or more structural forms. More recent findings suggest that there are some protein which are fully functional also when they are in an unstructured form [66]. In addition, in standard condition of temperature ($\sim$ 300 K), it is necessary to consider also the thermal noise, which produces a vibration around the protein structural conformation [64].

2.1.4 *The primary structure*

Protein primary structure is the linear sequence of amino acids in a peptide or protein. The peptide bonds, each of them holding 6 atoms, that link two consequent amino acids in a protein's chain is rigid and planar, preventing the peptide bond from rotating. The alpha carbon ($C\alpha$) in the center of each amino acid is held in the main chain by two rotatable bonds. The dihedral angles of these bonds are called φ and ψ [64].

2.1.5 *The secondary structure*

The secondary structure strongly reflects the local conformation of a protein. The secondary structure often follows common regular folding patterns of the polypeptide backbone. Only a few types of this

patterns are particularly stable and occur widely in a great variety of proteins: the most important are α helix and β conformations (β sheets and β turns, Fig.2.3).

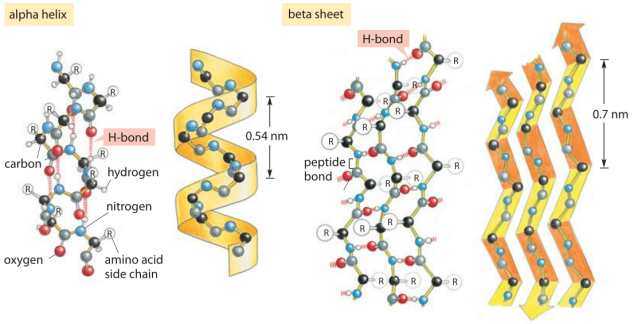

Figure 2.3.: The two most common type configurations of proteins secondary structure: α helix and β sheets.(Fig. adapted from Molecular Biology of the Cell (2002), 4th ed.)

α helix

The simplest, and also the most common, arrangement the polypeptide chain can assume, according to the presence of its rigid peptide bonds, is a helical structure, called the α helix. In this structure the polypeptide backbone is tightly wound around an imaginary axis drawn longitudinally through the middle of the helix, and the R groups of the amino acid residues are arranging outside the helical backbone. The repeating unit is a single turn of the helix, which extends about 5.4 Å along the long axis. The amino acid residues included in each helical turn are about 3.6 and the helical twist is always right-handed. More generally, about one-fourth of all amino acid residues in polypeptides are organized in α helices, while the exact fraction varying greatly from one protein to another.

The α helix are the most common structures because they make optimal use of internal hydrogen bonds. The whole helix structure is stabilized by an hydrogen bond between the hydrogen atom attached to the electronegative nitrogen atom and the electronegative carbonyl oxygen atom of the fourth amino acid of peptide bond. Within the α helix, every C=O and NH involved in a peptide bond (except those close to each end of the helix) participates in such hydrogen bonding. Each successive turn of the helix is held to adjacent turns by three to four hydrogen bonds: they give the entire helical structure considerable stability [65].

β sheets and β turns

The second type of repetitive structure is called the β conformation: in this case the backbone of the polypeptide chain is extended into a zigzag rather than helical structure. The zigzag polypeptide chains

can be arranged side by side to form a structure resembling a series of pleats. In this arrangement, called β sheet, hydrogen bonds are formed between adjacent segments of polypeptide chain. The R groups of adjacent amino acids protrude from the zigzag structure in opposite directions upward and downward and they can create an alternating pattern in the plane of the βsheet. The two adjacent polypeptide chains in β sheets can be either parallel (having the same amino-to-carboxyl orientation) or antiparallel (in this latter case, the structure is called β sandwich). The structures are somewhat similar, although the repeat period is shorter for the parallel conformation (6.5 Å, versus 7 Å for antiparallel) and the hydrogen bonding patterns are different. This structure is particularly abundant in the sovra-molecular assembly, as is one of the main conformational arrangement for amiloydal proteins when organized in fibrils.

In globular proteins, which have a compact folded structure, nearly one-third of the amino acid residues are in turns or loops where the polypeptide chain reverses direction. These are the connecting elements that link successive runs of α helix or β conformation. Particularly common are β turns that connect the ends of two adjacent segments of an antiparallel β sheet. The structure is a 180° turn involving four amino acid residues. β turns can be easily found near the surface of a protein, where the peptide groups of the central two amino acid residues in the turn can form hydrogen-bond with water. Considerably less common is the γ turn, a three residue turn with a hydrogen bond between the first and third residues [65].

2.1.6 *Tertiary structure and protein conformation*

The tertiary structure of a protein and the spatial arrangement of its atoms in 3-dimensional space is called *conformation*. This possible conformation include any structural state of the protein which can be arranged without breaking covalent bonds. A change of this conformation could be done also by the rotation around single bonds. It's clear that a protein could achieve numerous conformations, considering that it contains hundreds of single bonds. Despite this, one or (more commonly) a few conformations generally predominate under biological conditions. The need for multiple stable conformations reflects the changes that must occur in most proteins as they bind to other proteins or molecules. The protein's conformations occurring under a given set of environmental conditions are usually the ones that are thermodynamically the most stable, so the ones that have the lowest *Gibbs free energy* (**G**). Proteins in any of their functional, folded conformations are called **native proteins** [65].

2.1.7 *Weak Interactions stabilize the protein conformation*

Talking about the protein structure, the term stability under certain conditions could be defined as the tendency of the protein to maintain its native conformation.

Native proteins are only marginally stable: the ΔG which separates the folded and unfolded states in a typical protein under physiological condition belongs to the range of only 20 to 65 kJ/mol. A given polypeptide chain can theoretically assume countless different conformations, and as a result the unfolded state of a protein is characterized by a high degree of conformational entropy. This entropy, and the hydrogen-bonding interactions of many R- groups in the polypeptide chain with solvent (e.g. water), tends to maintain the protein in its unfolded state. The chemical interactions that counteract these effects and stabilize the native conformation include disulfide bonds and the weak (noncovalent) interactions as hydrogen bonds and hydrophobic and ionic interactions. Focus on these weak interactions is especially important to understand how polypeptide chains fold into specific secondary and tertiary structures, and how they combine with other polypeptides to build up quaternary structures. In a typical protein, about 200 to 460 kJ/mol are required to break a single covalent bond, whereas weak interactions can be disrupted by 4 to 30 kJ/mol. Individual covalent bonds that contribute to the native conformations of proteins are clearly much stronger than individual weak interactions. Yet, because they are so numerous, weak interactions predominate as the stabilizing force in protein structure. In general, the protein conformation with the lowest Gibbs free energy (the native protein conformation) is the one with the maximum number of weak interactions (Fig.2.4). In addition, the stability of a protein is not simply the sum of the free energies of formation of the many weak interactions within it. Every hydrogen-bonding group in a folded polypeptide chain was hydrogen-bonded to water prior to folding, and for every hydrogen bond formed in a protein, an hydrogen bond (of similar strength) between the same group and water was broken. The net stability contributes by a given weak interaction. The difference in free energies of the folded and unfolded states may be close to zero. So it is necessary to look elsewhere to find the reason why the native conformation of a protein is favoured compared to the others.

A partially solution can be find by thinking about the particular properties of the water. Pure water contains a network of hydrogen-bonded H_2O molecules. No other molecule has the hydrogen-bonding potential of water. When a hydrophobic molecule is surrounded by water, the optimal arrangement of hydrogen bonds results in a highly structured shell, also called *solvation layer*. The increased order of the water molecules in the solvation layer correlates

with an unfavorable decrease in the entropy of the water. However, when nonpolar groups of the polypeptides are clustered together, there is a decrease in the extent of the solvation layer because each group is partially hidden inside the protein. The result is a favorable increase in entropy that is the major thermodynamic driving force for the association of hydrophobic groups in aqueous solution. At the same time, hydrophobic amino acid side chains tend to be clustered in the protein's interior, hidden from water. The formation of hydrogen bonds and ionic interactions in a polypeptide is driven by the same entropic effect. Proteins polar groups can establish hydrogen bonds with water: so they are generally soluble in water. However, the number of hydrogen bonds per unit mass is generally greater for pure water than for any other liquid or solution: so there are limits to the solubility of even the most polar molecules. Therefore, the solvatations shell of structured water could also form to some extent around polar molecules. Even if the energy of formation of an hydrogen bond between two different polar groups of a macromolecule is canceled out by the elimination of the ionic interactions between the same groups and water, the release of water provides an entropic driving force for folding. The main part of the net change in Gibbs free energy that occurs when weak interactions are formed within a protein is derived from the increased entropy in the surrounding aqueous solution resulting from the burial of hydrophobic surfaces. This more than counterbalances the large loss of conformational entropy as a polypeptide is constrained into a single folded conformation [64].

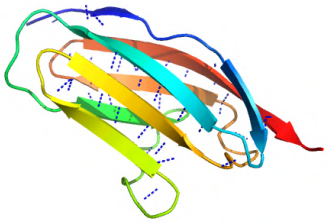

Figure 2.4.: The structure of a single molecule of titin I27. The dot blue lines represent the intramolecules hydrogen bonds (Fig. obtained by PyMol)

Hydrophobic interactions are clearly important in stabilizing a protein conformation; the interior of a protein is generally a densely packed core of hydrophobic amino acid side chains.

Even though one hydrogen bond seems to have a little importance in the stability of a native structure, the presence of hydrogen bonding and charged groups without partners in the hydrophobic core of the protein can be so destabilizing that conformations containing these groups are often thermodynamically untenable. The favorable free energy change realized by combining such a group with a part-

ner in the surrounding solution can be greater than the difference in free energy between the folded and unfolded states. Additionally, it is known that hydrogen bonds between groups in proteins form cooperatively. The contribution of hydrogen bonds and other noncovalent interactions to the stabilization of protein conformation is still being evaluated. One of the other concomitant factors that contribute in the protein stabilization is the interaction of opposite charged groups which forms an ion pair (*salt bridge*).

Most of the structural patterns of proteins reflect two simple rules: (1) hydrophobic residues are largely buried in the protein interior, away from water; and (2) the number of hydrogen bonds within the protein is maximized. Insoluble proteins and proteins within membranes follow somewhat different rules because of their function or their environment, but weak interactions are still critical structural elements [64].

2.1.8 *Protein folding pathway and denaturation*

Each protein begins its existence as a linear sequence of amino acids residues. This polypeptide must fold during and following synthesis to reach its native conformation. This native protein conformation is only marginally stable, indeed a modest change in the environment can involve different structural changes that affect the protein conformation, leading to a misfolded event.

In living cells, proteins are composed by a great numbers of aminoacids: during the birth of a new protein, aminoacids are joined in a very high rate. If a 100 amino acids protein is considered, and assuming that each residue could take up 2 different space directions on average, 2^{100} different conformation for the polypeptide are equally possible. Assuming that the protein folds itself spontaneously by a random process in which it tries out all possible conformations around every single bond until it finds its native, biologically active form and assuming that each conformation is sampled in the shortest possible time (1-1.3 s) it would take about 10^{10} years to sample all possible conformations, which is more than the age of the Universe.

Thus protein folding cannot be a completely random, trial-and-error process: there must be some shortcuts. This problem was first pointed out by Levinthal in 1968 and it is sometimes called **Levinthal's paradox**. On the other side, the folding pathway of a large polypeptide chain is unquestionably complicated, and not all the principles that guide the process have been worked out. In one of the proposed models, the folding process is envisioned as hierarchical. Local secondary structures were formed first: some amino acids folds into α helix or β sheets, then they are followed by longer-range interactions in order to create supersecondary structure: it seems a

kind of cooperative process that continues until the entire polypeptide is folded, usually in a cooperative way.

An alternative model also exists, in which the folding is driven by hydrophobic interactions among nonpolar residues. In this process, the protein first collapses into a compact state, sometimes called molten globule state, which may have an high content of secondary structure, but the real position of each amino-acid is still flexible.

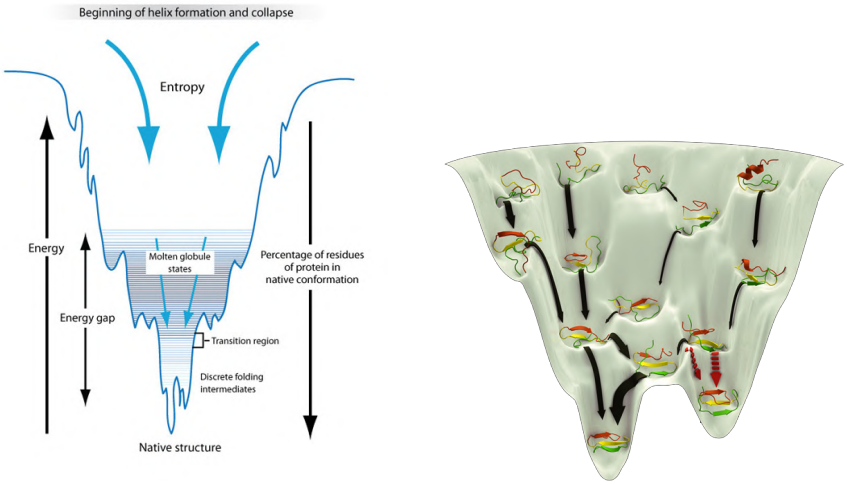

Figure 2.5.: Thermodynamics of protein folding depicted as a free-energy funnel. As folding progresses, the thermodynamic path down the funnel reduces the number of present states, increases the fraction of protein in the native conformation and decreases the free energy (Fig. taken from ®Annenberg Learner and ®Thomas Splettstoesser).

The rules followed by the process of protein folding are not clearly determined: most proteins probably fold by a process that is the mixture of both models. The surprising thing is that, instead of following a single and unique pathway, a population of peptide molecules may take a variety of routes to reach the same end point. During this process the number of different partly folded conformational species decreases as folding reaches completion. Thermodynamically, the folding process can be seen as a kind of free-energy funnel as shown in Fig.2.5. The unfolded states have a high degree of conformational entropy and relatively high free energy. As folding proceeds, the narrowing of the funnel represents a decrease in the number of conformational species present. Small depressions along the sides of the free-energy funnel are semistable intermediates that can briefly slow the folding process, or "capture" a protein which will not reach its stable conformation. At the bottom of the funnel, an ensemble of folding intermediates has been reduced to a single native conformation. This process is highly not linear and it can cause defects in protein folding, generating a misfolded that may be the molecular

basis for a wide range of human genetic disorders. In nature, the folding pathway is often assisted by molecular *chaperones*, proteins that help the conformational folding or unfolding of other macromolecular structures [65]. Each protein has regions of high and low stability: also small differences in regions of low stability could allow a protein to alter its conformation between two or more states. This little variations in the stability of regions within a given protein are often essential to protein function. A correct protein folding is essential to have a biologically working protein, but there could be conditions in which the protein loses its native conformation. Indeed, protein structures have evolved to work in particular cellular environments and environmental conditions. The change of this conditions could be so important to involve a loss of 3-dimensional structure that causes loss of function. This process is called *denaturation*. The denatured state is not necessary a unique conformation: under most conditions, denatured proteins exist in a set of partially folded states. Most proteins can be denatured by heat, which affects the weak interactions in a protein in a very complex way. If the temperature is increased slowly, the protein conformation generally does not change until the reach of a narrow temperature range. The abruptness of the change suggests that unfolding is a cooperative process: loss of structure in one part of the protein destabilizes other parts. Proteins can be denaturated also by extremes of pH, by denaturant as urea or by mechanical strength. Each of these denaturing agents represents a relatively mild treatment: no covalent bonds in the polypeptide chain are broken. Because of each denaturing agent causes different alteration in proteins conformation, involving different processes as the disrupting of hydrophobic interactions or hydrogen binding, the denatured states obtained with these various treatments need not be equivalent [65, 64].

2.2 THE MECHANICAL UNFOLDING OF SINGLE PROTEINS

The understanding of the mechanisms driving the folding an the unfolding of a protein is still a challenge. Different and complementary informations could be provided by several experiments, which may include also ensemble-average techniques, such as NMR, small-angle X-ray scattering and other spectroscopy techniques and also computer simulations which provide essential informations that are otherwise inaccessible because of the technical limitations.

Even if in the past the chemical (or thermal) unfolding has been compared to the mechanical unfolding [67] and the unfolding rate constants and the height of the energetic barrier had comparable values [67, 68], some careful distinctions must be done. First of all, thinking about the final state of the protein in denaturant conditions or after a mechanical force (Fig.2.6), it is evident that they have completely

different spatial arrangements: random coil for the first and extended status for the latter.

The chemical unfolded state usually follows the Flory's scaling law [69, 70] for a random coil in a good solvent whereas the force unfolded state is well described by the Worm Like Chain model (WLC) (see Appendix D.3).

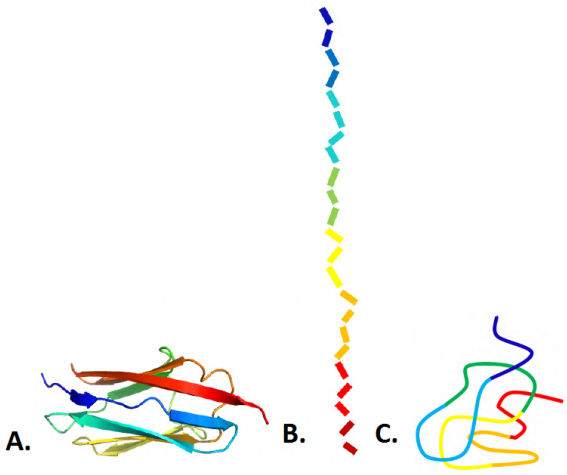

Figure 2.6.: (A.) Native structure of a protein, (B.) Protein force-unfolded configuration and (C.) Chemical unfolded state.

In order to compare the two unfolding processes, that evidently involve different final states, it is necessary to investigate how the energetic landscape changes while an external force is applied to the protein. In the first approximation, the folding-unfolding transition may be described as a two state process, separated by an energetic barrier (Fig.2.7, black line). When a protein is in its native folded state, its has already reached a free energy minimum and it will not unfold until an external force in the mechanical unfolding, or the increase of the temperature in the thermal one, allows it to overcome the energy barrier (ΔG^{TS-F} in Fig.2.7).

The external factor, which lowers the energetic barrier, may be the increase of the temperature or the addition of a denaturant. Another way to break the hydrogen bonds, which are responsible of the stable protein conformation, is to apply a mechanical force and lower the energetic barrier. In this case the protein is stretched along its reaction coordinate. This stretching force allows the energy landscape to tilt and, consequently, lowers the height of both energy barrier and the energy minimum of the unfolded state. So, if a force is applied to a protein, the unfolded state becomes more energetically favorable and more populated.

Several proteins, in nature also present in the human body, are continuously subjected to mechanical force, for examples proteins involved in the muscular activity or into the transport across the cellu-

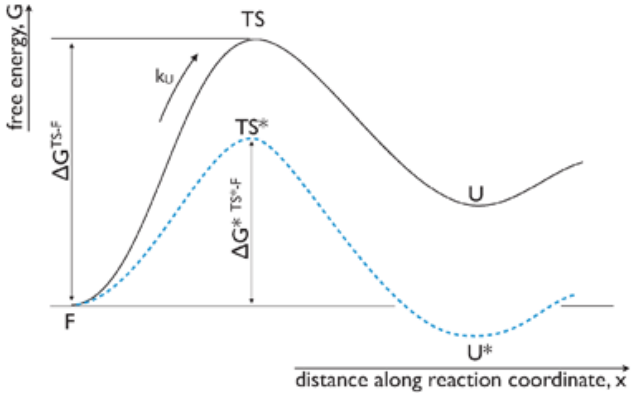

Figure 2.7.: Unaltered energy landscape for a protein (black line) and after the application of an external force (dot blue line) (Fig. taken from [81]).

lar membrane. In principle, these proteins must be able to unfold because of a mechanical force and then refold in their functional state. The study of their mechanical unfolding process is impossible by using thermal or chemical denaturation: while it could be well described by using single-molecules force spectroscopy techniques. In this case, by applying an external force, a single protein unfolding event can be observed in an environment as close to the physiological conditions as possible. Previous studies have shown that the unfolding rate constant extrapolated at zero force qualitatively matches the unfolding rate constant in absence of mechanical force, i.e. during chemical or thermal denaturation [67]. Despite this, it is important to clarify that there is still a structural differences between the unfolded states in both thermal and mechanical unfolding (Fig.2.6). Indeed, it has been demonstrated by MD simulations [71] or by single molecule experiments [72] that the mechanical and chemical unfolded proteins exhibit some differences in their local chains structure.

2.2.1 *The use of engineered polyproteins in Force spectroscopy*

Even if it is theoretically possible to pull a single protein anchored between the AFM cantilever tip and the substrate, how it could be clearly distinguished from a-specific adhesion or other molecules in the buffer was one of the main problems related to this single molecule technique. Indeed, when the tip is pressed against a layer of proteins, many molecules may interact with the tip, provoking the bending of the cantilever. The first problem is that, even if an unfolding event is detected, an evidence is needed to differentiate true unfolding events from spurious interactions: indeed only rarely the molecule of interest is individually picked up. More often the interactions will involve a fragment of protein, or more than one molecules. The second problem to be faced is that the proteins are relatively

short: they may be hidden by other kind of nonspecific interactions [14].

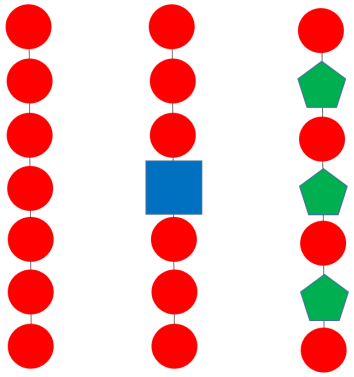

Figure 2.8.: Three different kinds of chimeric polyproteins: left, homomeric polyprotein, formed by identical modules, center: the protein object of study (blue square) inserted in two series of a standard proteins used as a fingerprint and right: polyprotein formed by two alternated different modules (red circles and green pentagons).

The use of engineered multidomains proteins overcomes this limit: the repetition of the same modulus gives a force spectroscopy trace that has a clear sawtooth pattern that could be easily distinguished from any other curve [73]. However, modular proteins with different modules, which may occur in nature, have an heterogeneous population regarding the length, force stability and sequence. The construction of homomeric proteins, which are made by $4-10$ identical domains, allows the study of a specific protein. By using a recombinant DNA, it is possible to construct a polyprotein with the desired characteristics, also by including a protein of interest in a more *standard* protein chain (see Chapter 5). Each single protein is linked to its neighbour by a short linker (usually made of 2-3 aa), whose length must be considerate during the Force Extension experiment in order to do not overestimate the contour length of the single protein.

Taking advantages of recombinant techniques, the protein under investigation can be insert in a chain of several repetition of the same domain or can be linked to other proteins that are used only as fingerprint. Alternatively the two different kind of proteins can be alternated (Fig.2.8).

By using polyproteins in force spectroscopy in the early 2000s, some points have been clearly assumed as evidence of real unfolding of protein and not to the detachment of denatured proteins absorbed onto the surface:

- The unfolding rate extrapolated at zero force is comparable to the one obtained by chemical denaturation experiments [67, 75]. This provides strong evidence that saw-tooth patterns represents true protein unfolding events.

- During the stretching by AFM, the distance between peaks is reproducible and corresponds to the expected values based on NRM studies or MD simulation. Different experiments [76] clearly show that the spacer between the modules are extended prior to unfolding.

- Force extension curves of a heterodimeric protein $(I27 - I28)_4$ (Fig.2.9B-C.) show saw-tooth pattern where the force peaks are divided into two different clusters: the weak forces that occur at the beginning of the trace represent the unfolding of I27 domains. If the force peaks were due to the absorption, an alternate pattern of force peaks would be expected [75].

- When a protein with a very big variety of modules (as titin) is stretched, the force peaks are ordered form weakest to strongest [77].

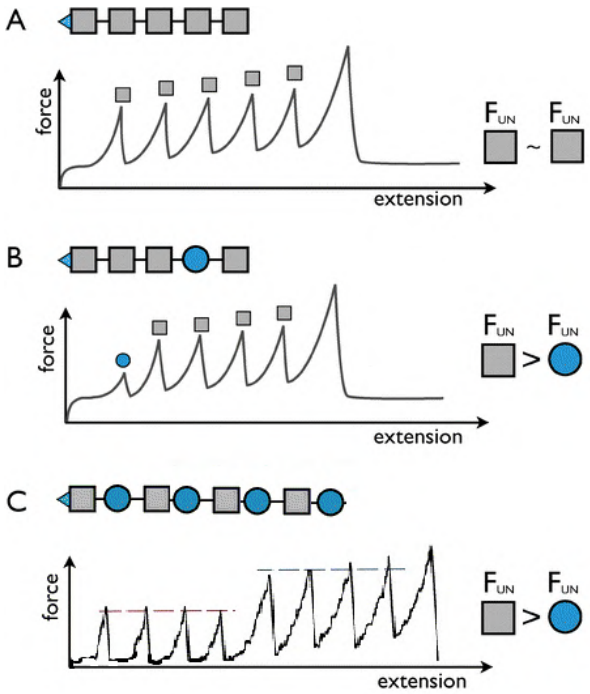

Figure 2.9.: Three different saw-tooth patterns of chimeric polyproteins: (A) homomeric polyproteins, (B) protein object of study (blue circle) inserted in two series of a standard proteins used as a fingerprint and (C) polyprotein formed by two alternated different modules. Figure adapted from [14] and [75].

2.2.2 *A strongly mechanically stable protein: the I27*

The giant elastic protein titin is the responsible for the human heart contraction. It is the direct regulator of the quantity of blood filling

in the left ventricle during diastole and problems in titin mechanical functions are directly related to the insurgence of several heart diseases [78]. By the control of the passive elasticity of muscles, mutations of this protein play a key role in dilated cardiomiopathies and centronuclear myophaty. Studies about titin post-translational modification show that it regulates also the tissue elasticity [79].

The structure of Titin is very complex (Fig.2.10), as it is formed by modules of proteins that are very different to each others. It is made up of immunoglobulin-like (Ig) and fibronectin type III (fnIII) domains with additional less structured regions such as the N2B and PEVK region. The half I-band connects the Z-disk and the A-band and is composed of distinct regions, referred to as the proximal immunoglobulin-like region, N2B and PEVK region, and distal immunoglobulin-like region. The most investigated sub-sections are located in the latter region. Since it is physiologically subjected to mechanostructural modifications during heart contraction, titin is a perfect protein to be investigated in presence of an external force in near-physiologic conditions. Even if titin has a great number of modules, it has been shown that they fold and refold independently. Force studies may be done both on the entire protein and on one or few of its modules.

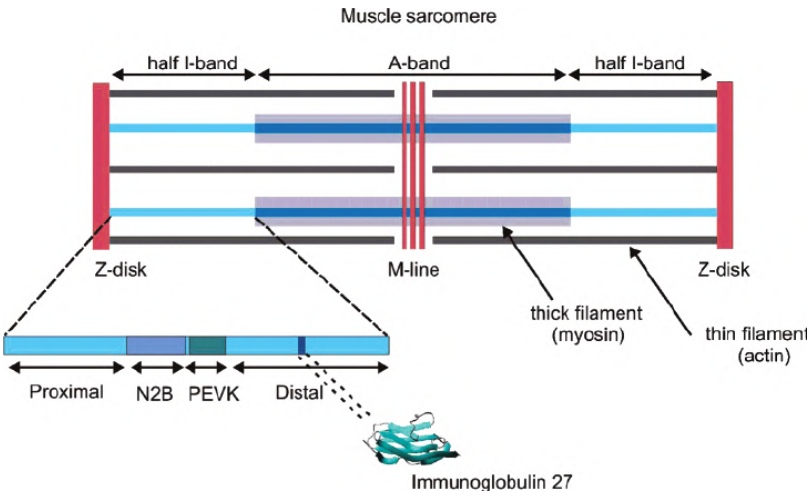

Figure 2.10.: Schematic diagram showing the position of titin in the sarcomere. The immunoglobulin-like protein I27 is located in the I-band. Figure taken from [81].

One of the most studied modules is titin I27 (in the old nomenclature, I91), which is an Immunoglobuline-like domain, located in the distal region of the half I-band. This protein shows an high resistance to unfolding when subjected to an external force. It is extremely stable, and its characteristics, as well as its length and its tertiary structure, are well known. For this reason it is a perfect protein for

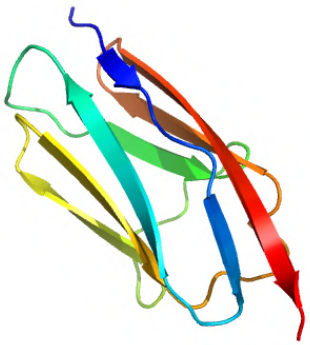

Figure 2.11.: Structure of titin I27, showing the position of its β sheets (image obtain by PyMol).

calibration in force spectroscopy studies and it is used as a fingerprint when chimeric[4] polyproteins are constructed.

The I27 modulus is a β-sandwich[5] protein, made of 89 amino-acid, and has a total contour length of $\Delta L_c = 28.1 \pm 0.2$nm [73] and a molecular weight of $\sim$ 13kDa [80]. It has seven β sheets: this structure is common to almost each immunoglobulin segment and it is responsible for the elastic proprieties of the Titin. It is also very stable at neutral pH ($T_{melting} \sim 60^{o}$C) so, it is not subjected to degradation or denaturation in the typical conditions of force experiments.

2.3 STRUCTURED PROTEINS AND IDPS

The proteins can exist in at least three different conformational states: the ordered (or folded) state, the molten globule or the random coil [82], called *the protein trinity*. In the last year, this trinity is described more as a continuum of conformations ranging from structured to fully unstructured forms [83]. In the literature exist a multitude of example demonstrating that the native protein conformation could correspond to any of these three states an its function could arise from any states or transition between them. For examples, the muscular protein titin is able to refold itself among multiple mechanical denaturations, and its work is directly translated in the muscular stretching. Its structure is a continuously alternating transition between an ordered state and a random coil. A relevant number of proteins that undergo to chemical modification, as hyperacetylation, show a less rigid conformation. A structural rearrangement is essential also in critical cell signalling events, as, for example, the calcineurin protein

4 Chimeric proteins are fusion of two or more proteins which are produced as a single polypeptide chain. In SMFS, chimeric proteins are widely employed for inserting a new protein, the real objected of the study, in a chain of other well-known modules used as reference (Fig.2.9 B)

5 β-sandwich domains are characterized by two opposing antiparallel β-sheets

bound to a calmodulin: in this process is essential the presence of a unstructured region near the α-helix calmodulin target. It is quite frequent that a protein can present simultaneously more than one conformational state and/or could be divided into subregion presenting different orders of structure. The unstructured region is always linked to the presence of linker, molecular recognition domains, entropic springs and inhibitors domains. It has been showed that the amino acid sequence determine the native state of a protein: thus it helps, among different techniques, to predict the structural behaviour of a protein. It has been reported that unstructured polypeptide segments with less than 50 residues are quite common among functional proteins, and that, moreover, $\sim$ 30% of eukaryotic proteome encodes for totally or partially Unstructured Proteins [84], and this percentage increases as increase the complexity of the organism. These particular proteins (here referred as Intrisically Disordered Proteins (IDP) or somewhere called Intrinsically Unfolded Proteins (IUP) or Natively Unfolded Proteins (NUP)) show an unstructured conformation in the absence of a molecular partner, moreover their structure is not simply a Random Coil, but it is more similar to an high flexible chain that can undergo to several conformation rearrangements in the presence of external agents or during a time evolution [85].

Nevertheless, these IDPs are fully involved into several biological functions, which are somehow intimately linked to their particular structure. Moreover, some of these proteins shows different structural plasticity and well defined conformations and it has been proposed that this mechanism driving the protein evolution through the modern protein folding pathway, raising the IDPs as the *protein ancestors*. More recent findings, basing on the abundant number of different IDPs in the most complex organism, that the unstructured conformation could somehow be considered as the final evolute form [86, 87].

2.3.1 *Structure and function of the intrinsically disordered proteins*

The discovery of the IDPs is not recent, as they were first described in [84] during the last years of 1970s. Moreover, these proteins are hard to been purified directly from the cells, due to their low copies number, and they are subject to a fast degradation. Nevertheless, only in the last 20 years the research focused on describing the IDPs: they were particularly difficult to be characterised by X-ray Crystallography, which is considered as the most powerful technique for describing protein structure: in fact the disordered regions are visualized as zone lacking electron density. The IDPs started to be investigated after the development of more sensitive techniques, as Nuclear Magnetic Resonance or Mass Spectrometry, still suffering to the limitations of bulk measurements, like the averaging and the complexity of the identification of different conformational structures. Finally, in

the last ten years, single molecule approaches have been employed also to map the structural complexity of disordered proteins [88, 89] and to probe their molecular structure.

The function of the IDPs has not been fully explained yet, but it is clear that the protein disorder must have a functional rationale. Indeed, it has been proposed that the flexibility of the IDPs along with their ability to bind an extended number of different ligands and their big intermolecular interface maintaining a relatively small size are unique features. Usually, inside an healthy cell, proteins which fail to fold undergo to an ubiquitin mediated degradation, but this is not happening for IDPs, suggesting that their fully functional form may be partially structured when bound to its partner or that their rapid degradation leads to a functional dynamic regulation [90, 91].

The IDPs hardly could be described by using the protein trinity frame [82], as, even when they display an high elongation, the IDPs still displays tendency to form structured elements or cluster of hydrophobic residues: these features are extremely important to achieve the binging/interaction with ligands. For example, the α-synuclein protein tend to form a partial α-helix structure when it is bound to a lipidic micelle or to a surfactant [92, 93, 94]. It has been proposed that the disorder-to-order transition, occurring when an IDP ordinate its conformation when bound to ligands, has an entropic cost: to be energetically convenient, its thermodynamic driving force for the binding reaction should be consequently a favourable enthalpic contribution. This binding-coupling combination could be described as highly specific but of low affinity. More recently, this concept had been overcome by Borgia *et al.* [96] in which an ultrahigh-affinity protein complex between two intrinsically disordered proteins was described. In particular, the histone H1 and its nuclear chaperone prothymosin-α had found to associate in a complex with picomolar affinity, but fully retaining their structural disorder, long-range flexibility and highly dynamic character. This unexpected interaction mechanism can be explained by the large opposite net charge of these two proteins, without requiring structurally defined binding sites. In general, the energy landscape of IDPs is more lable than the structured proteins, little microenvironmental changes, like the presence of ligand, increase or decrease of pH or ionic strength, the presence of metals or alcohols, could change the IDP energy funnel, driving the proteins to different structural states. In most cases, the ID proteins are bound to a partner in physiological state, showing a structured or partially structured configuration. When these proteins are purified for *in vitro* measurement the IDPs start showing their disordered proprieties [95]. It has been demonstrated that the cellular crowding is not sufficient to induce the IDP folding [97], also by considering the evidence that the extracellular IDP, physiologically

living in a less-crowded environment, show the same unstructured conformation [98].

Furthermore, some IDPs, such as the prothymosin-α, adopt unstructured conformations even in absence of the binding partner in cells as shown in [99] by single molecule fluorescent measurements in living cells. By using nuclear magnetic resonance and electron paramagnetic resonance spectroscopy in living mammalian cells, Selenko and co-authors showed that the disordered nature of monomeric α-synuclein protein is preserved in both neuronal and non-neuronal cells [100].

The human α-synuclein (AS) is a small protein constituted by 140 amino acids which belongs to the synucleins family [101], and it is encoded by the SNCA gene. It is an acidic, soluble and heat stable protein, with an observed molecular weight of approximately 14.5 kDa [102]. AS does not acquire a secondary structure when purified from heterologous expression system, for this reason it has always been considered an intrinsically disordered protein [103]. The primary structure of AS is usually divided into three different parts: N-terminus, which can acquire a secondary structure, non-β amyloid component (NAC domain) and C-terminus, always disordered [104] (see Fig.2.12). The amino-terminal sequence of AS is constituted by the first 60 amino acids and is almost entirely composed of variants of an imperfect 11 amino acid repeat [105] (Fig.2.12). It can form, along with the NAC region, an amphipatic α-helix responsible for AS interaction with lipidic membranes [92, 106]. The single point mutations A30P, E46K and A53T [107, 108, 109] which are marker for familial form of Parkinson's Disease (PD) are all located in this region, accounting for variation in protein-membrane interaction or in protein aggregation propensity. Recently, other two mutations have been found in this region: H50Q [110] and G51D [111] (Fig.2.12).

The second region, the so called NAC portion, ranges from residue 61 to residue 95. This segment is the most hydrophobic portion of the protein and it is believed to be responsible for the primary intramolecular interactions that cause AS misfolding and aggregation. The NAC is the second major component of amyloid plaques in the brains of Alzheimer's Disease (AD) patients [103], acquiring a β-sheet structure when AS forms amyloid fibrils [112]. Recently, the functional mutation E83A has been discovered in this core region of AS, suggesting a possible role for the E83 residue in preventing protein aggregation. As a matter of fact, the E83A mutant has been reported to form cross β-structures involved in fibrils formation and aggregation [113].

The last portion of the AS protein constitutes the C-terminus, which includes 14 acidic residues and does not acquire a defined secondary structure in solution. However, it has been shown to regulate fibril formation, both when AS is bound to the membranes and when it forms amyloid fibrils [114].

AS is a good conserved protein among many different types of organism. The human AS gene (SNCA) encodes three different isoforms by alternative splicing [115, 116]. The major and most abundant isoform is derived from the complete protein transcript, and is, indeed, 140 amino acids in length. The two other AS isoforms, SNCA-126 and SNCA-112, result from in-frame deletions of exons 3 (residues 41-54) and 5 (residues 103-130), respectively. The different

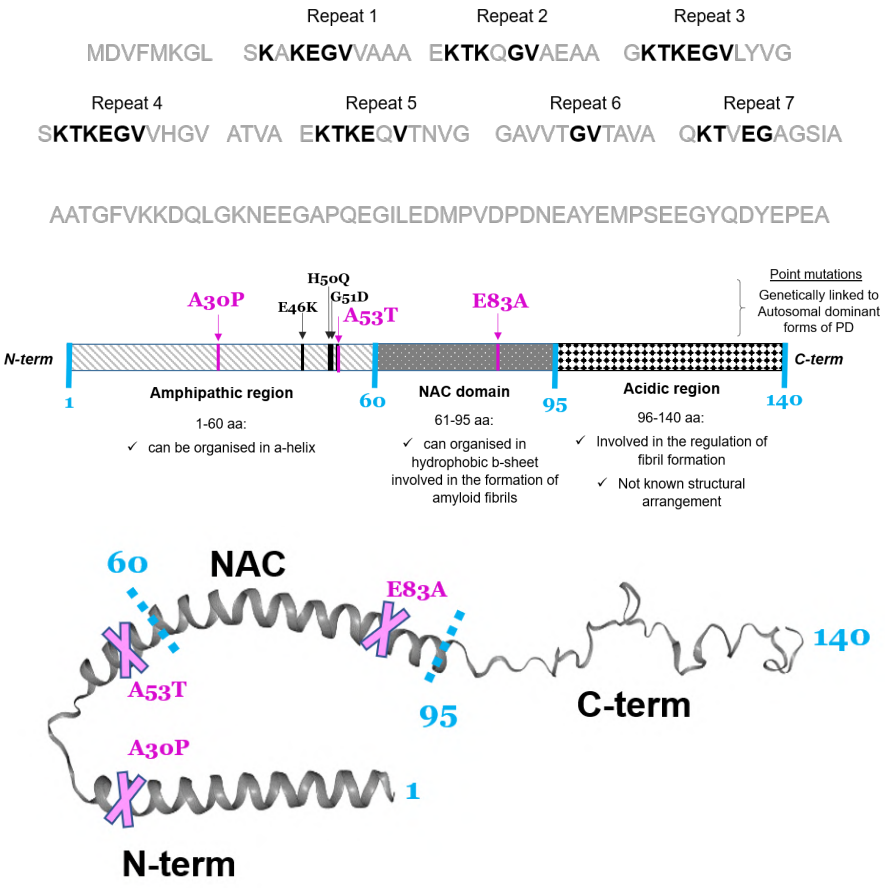

Figure 2.12.: **Primary structure of AS.** (Top) The imperfect 11 amino acid repeats are emphasized, and the more conserved residues are represented in boldface. In the AS sequence (Middle), the three different regions that compose the protein are indicated. The N-terminal amphipatic region contains most of the repeats and the five point mutations linked to autosomal dominant early-onset PD (A30P, E46K, H50Q, G51D, A53T). The central region (NAC), which encompasses the most hydrophobic residues, promotes aggregation. The functional point mutation that was found in this region, E83A, is indicated. Finally, the acidic C-terminal portion of the protein that was shown to have a role in decreasing protein aggregation. In pink the three mutations discussed in this Thesis are highlighted. Figure inspired by [105].

functions of the three isoforms are not totally understood, indeed the majority of the studies have focussed on the role of the full length protein.

2.4.1 *Conformations and Structural Organisation*

Several studies concluded that, differently from the *in vivo* findings, the purified AS forms a natively unfolded structure in solution [95].

These conclusions were supported by isolation of the AS protein from several different sources (brain tissue, mammalian and bacterial cells) [117]. Additional studies proposed that the AS is present in a native tetrameric structure [118], in an oligomeric tertiary structure [119] or α helical octamers [120].

The debate on AS native structure is not closed yet, but the current picture depict an AS capable to exist into a wide range of different conformations depending on the environment, binding partners or membrane bilayers.

2.4.2 *Physiological Function*

The AS is highly expressed in the brain (representing up to 1% of protein content in neuronal cytosol) and is the most abundant component of the Lewy body (LB), proteins aggregates found in dopaminergic neurons, which are typically of the Parkinson's disease (PD) and other neurodegenerative disordered, called synucleopaties. Despite the massive presence of this proteins, its physiological function has not been clearly elucidated yet [121, 122]. The majority of the AS, both the free and AS:plasma-membrane bound protein, vescicles or micelles, populates the presinatic-termini. Several studies suggested that the AS plays a role in endoplasmic reticulum-Golgi vesicle trafficking and synaptical function in the nervous system. It may also help regulate the release of dopamine (DA) in the brain. A considerably relevant portion of AS (15%) is, indeed, usually bound to membrane, and some studies suggested an antioxidant effect of AS. The flexible structure of the AS, and the *in vivo* presence of both unstructured components as well as more structured ones, seems to play a key role in the molecular recognition of their partners, such phospholipids, ligands and proteins. Indeed, AS has been described to specifically interact with several proteins involved in signal transduction, ubiquitin-proteasome system, vescicular recycling, synaptic functionality, regulation of oxidative stress and mitochondrial function [121]. The free AS in solution usually displays a lack of tertiary structure, while secondary-rich AS structures were found when bound to small micelles (α-helix) or when undergoes to pathological amyloid fibrillation (β-sheets).

2.4.3 *Role in Parkinson's Disease*

The initial link between AS and the synucleoinopathies was established by its isolation form insoluble aggregates in neuronal tissues of Alzhaimer's Disease (AD) patients. The AS was then associated with the Parkinson's Disease (PD) with the first discovery of A53T point mutation [109]. After the AD, the PD is one of the most common neurodegenerative disease in familiar form of the pathology, it

is mainly characterized by an extensive loss of dopaminergic neuron in the substantia nigra, leading to clinical insurgence of several movement disorders [123]. The main pathological features are the intraneuronal Lewy bodies (LB) and Lewy neurites, which are aggregated protein deposits, mainly consisted of AS. Here comes the considerable interest in investigate AS aggregation and its role in neurodegenerative diseases. Several studies have been conducted to elucidate the conformational structure of the AS involved in misfolding and oligomer formations. Studies focused on the detection the AS in the presinaptic compartiments suggest a potential involvement into the neurotransmission-related pathways [124].

2.4.4 *Fibrillation and Aggregations*

Due to the presence of aggregative deposits of AS into LBs and the PD-related single point mutations that could affect AS stability, functionality and aggregation propensity, a significant effort was put over the years into the study of the AS aggregation process *in vitro*.

While wild type AS (WT-AS) *in vitro* shows a prevalence of natively unfolded conformations, in several conditions the AS acquires a structure that has a high propensity to aggregate. It has been proposed that this propensity to aggregate could be strictly connected to the pathological misfolding. It is still not clear how the conversion from an unfolded form to a defined structure occurs, but it is probably involved in the initial process of oligomers formation. The majority of the studies are focused on two different conditions: in solution or at the lipidic membrane interface [126].

AS conformational changes were studied in solution by nuclear magnetic resonance (NMR) and molecular dynamics (MD) [127, 128, 129], single molecule fluorescence techniques (SMF) [130, 131] or after a phase-gas transfer by native mass spectrometry (native MS) [122]. All these studies identificated long range interactions between C-termini and the core of the AS protein (NAC region) and highlited a continuous rearrangements of the two termini, suggesting that the unfolded state of AS in solution is constituted by an ensemble of different, transient and dynamic conformations. Some of these conformations can be, in particular conditions, prone to establish transient or stable bounds with other AS molecules, driving to protein oligomerization [105].

In the presence of lipid membranes, it was observed an increased level of AS aggregation propensity, indicating that AS α-helical structure directly converts into a β-sheet aggregation prone conformation, as it has been shown for other amyloidogenic proteins [126].

The first step of AS monomers self-interaction is mainly driven by hydrophobic interactions due to a hydrophobic stretch in the middle of AS NAC sequence (from residue 71 to 82), constituted by 12

amino acids (VTGVTAVAQKTV). After this, monomers start to aggregate and form oligomeric-like species. As soon as their concentration reaches a critical point, these oligomers may be rapidly converted into protofilaments, protofibrils and, finally, into amyloid fibrils [112, 132].

Recently, it has been proposed that the AS oligomeric species were the most toxic aggregation intermediate in the AS fibrillization pathway [133]. Indeed, it was observed that the AS PD-linked A53T and A30P mutants shared an accelerated oligomerization *in vitro*, but not an accelerated fibrillization, suggesting that an increased fibrillization rate is not directly linked to a higher cytotoxicity in PD [134]. Accordingly, also the E46K mutant, which shows an increased aggregation rate with respect to WT-AS, has been demonstrated to form soluble non fibrillar species [135].

The biophysical characterization of AS oligomeric species is particularly difficult, by the fact that they often form a very heterogeneous ensemble, constituted by transient species that are difficult to isolate and analyse [133]. However, several efforts were put in the study the oligomeric species [134, 135]. These studies focused on analysing the aggregation deposits obtained from the pathological mutants and found a heterogeneous ensemble of oligomers and protofibrils. Some of these soluble oligomers and protofibrils showed annular or elliptical shapes, probed by atomic force microscopy and electron microscopy, with diameters going from 4.5 to 55 nm. The estimation of the average number of AS monomers constituting these oligomers goes from 16 to more than 42 [136].

Several problems arose from the *in vitro* studies of these species: oligomers structures, morphological and functional characteristics are extremely dependent on experimental conditions employed both for their formation and their analysis. The final oligomeric picture shows a wide variability in the obtained structure: these oligomers, mainly obtained *in vitro*, are not always found *in vivo*. The most recent efforts concern mainly their characterization in a more physiological environment [105].

The structure of AS amyloid fibrils has been characterized by several complementary techniques, such AS AFM or SEM imaging and MD. The obtained data suggested that AS fibrils have a β-sheet core region (including not only the NAC domain), not exposed to the solvent and organized in an ordered in-register parallel structure with β-strands planes perpendicular to the fibril axis. 5 or 6 β-strands, turns and loops were identified within this core region. Moreover, N-terminus showed a more heterogeneous and less ordered structure, remaining still quite rigid, while C-terminus was described as flexible and completely unfolded [105].

2.4.5 *Interaction with ligands and point mutations*

AS fibrillization is strongly affected by various factors that can enhance or inhibit aggregation or rearrange the structural conformation of the protein. It has been shown, by AFM imaging and dynamic light-scattering, that fibrils elongation rate is directly proportional to protein concentration [138], which is in good agreement with the fact that SNCA gene triplication, responsible for rare and aggressive genetic PD forms, causes an increased dosage of AS [139].

Covalent chemical modifications of AS, such as nitration, oxidation and phosphorylation (S87 and S129), lead to variation in the rate and in the final products of the aggregation process. For instance, it has been demonstrated that the oxidation of AS drives its oligomerization, while the monomeric and dimeric nitrated AS promoted fibril formation [140].

A recent study has discovered that AS constitues a highly efficient target for mid-chain arginylation and is arginylated *in vivo* on two E residues, E46 and E83, that are implicated in familial cases of PD [141].

Divalent and trivalent metal ions (such Al^{3+}, Cu^{2+}, Fe^{3+} or Pb^{2+}) have been shown to strongly accelerate AS aggregation rate *in vitro*. The effect of these compounds suggests a critical role for electrostatic interactions in the aggregation rate modifications.

2.4.6 *AS interaction with EGCG*

Epigallocatechin-3-gallate (EGCG) is a polyphenol that belongs to the catechins family, naturally produced by the *Camellia sinensis* plant specie. The EGCG structure consists of four rings resulting from the esterification of epigallocatechin (EGC) with gallic acid. EGCG's numerous hydroxyl groups as well as its conjugated structure, ideal for electron delocalization, confer on this molecule important free radical scavenging properties [143].

It has been demonstrated to have a neuroprotective role in PD on the basis of its anti-oxidative, antiinflammatory, and fibril-destabilizing effects. It has been also shown to cross the BBB and to be well tolerated in humans. There are still some obstacles regarding its rapid metabolization in the liver, although several new drug delivery systems, such as encapsulation of EGCG in nanoparticles or nanolipidic capsules, have shown promising results for increasing the bioavailability of EGCG [142].

EGCG is capable of modulationg the misfolding of proteins by directly binding to unfolded polypeptide chains and forming unstructured, nontoxic oligomers intead of the ordered β-sheet-rich aggregates [143]. In the case of AS, EGCG not only inhibits the conformational transition from random coil to β-sheet confomers, but it has

been shown also to be capable to disaggregates the amyloid fibrils in a dose-dependent manner. Therefore, it could be speculated that when the AS is in an unfolded conformation, EGCG may block the β-sheet formation in the core amyloidogenic sequence, thus effectively inhibiting the self-assembly of AS fibrils [143].

Furthermore, all-atom MD simulations have confirmed that EGCG could to reduce the structural stability of AS fibrils and remodel them by converting the ordered structure into the disordered one. Three possible binding sites of EGCG on the surface of AS fibrils were found and it was shown that EGCG is capable of binding AS by hydrophobic interaction and H-bond interactions at the area rich in β-sheet secondary structures [144]. In figure 2.13 there is an exemplification of the interaction between EGCG and AS pentamer at the molecular level.

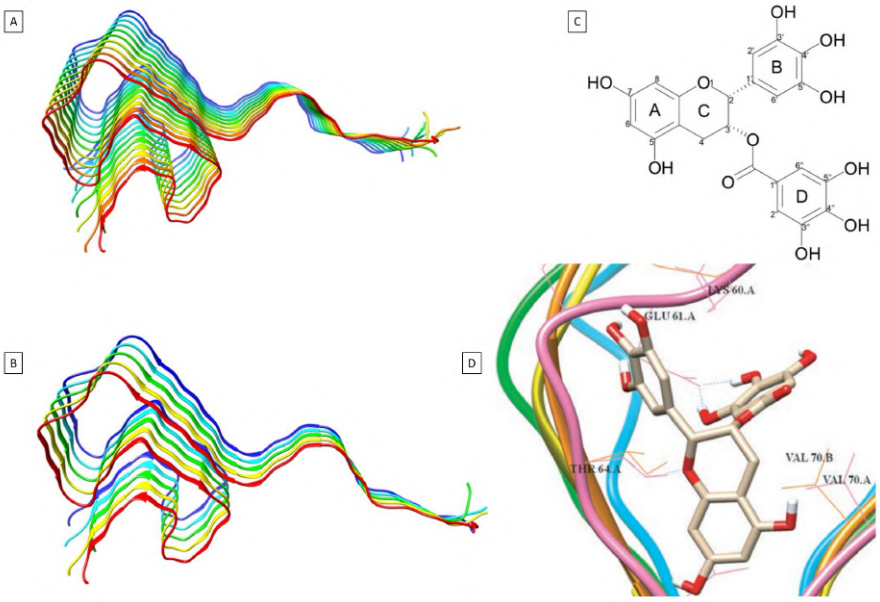

Figure 2.13.: A) Molecular models of AS decamer.
B) Molecular models of AS pentamer.
C) Molecular structure of EGCG.
D) An exemplfication of the static interaction mode of EGCG and S3 site of AS.
Figure adapted from [144].

2.4.7 *AS interaction with Dopamine*

A common pathway, involving DA-dependent oxidative stress, has been put forward to explain the death of dopaminergic neurons in PD. Defects in the sequestration of DA into synaptic vesicles in dopaminergic neurons from the *substantia nigra* could enable undesired DA:AS interactions and form toxic intermediates found in LBs [146]. It has

been reported that AS changes its structure in response to DA, or most likely to by-products of DA oxidation, adopting a conformation where its N- and C-termini become closer together, thus, inhibiting AS fibrils formation and leading to the accumulation of AS oligomeric species via an alternative folding pathway [145].

The DA binding site of AS is located on the C-terminal region of the protein at $_{125}$YEMPS$_{129}$ which is highly negatively charged. This region was implied in the modulation of DA induced aggregation. AS:DA interactions are further stabilized by long-range electrostatic interactions with E83 in the NAC region, as reported by NMR and MD simulations [147]. Therefore, in the case of the functional mutant E83A, the replacement of E with A should act in the opposite way. The E83A mutation in the NAC region is, indeed, known to strongly impair the ability of DA to inhibit AS aggregation. This could be due to the destabilisation of AS:DA interaction, by either hindering DA binding or by altering some properties of the NAC region. This means that both the C-terminus and the NAC region are necessary for the inhibition of AS fibrillization by DA [148].

Electrospray-ionization-ion mobility spectrometry-mass spectrometry (ESI-IMS-MS) experiments have shown that the DA molecules bind exclusively to the extended conformation of AS, while binding is not observed in the compact state of the protein [147].

This result was confirmed by a more recent work, in which the structural effects of the binding of EGCG and DA to AS were compared [149]. From this study, which is further discussed in part II, it is shown that EGCG-AS binding prefers a compact conformation, while DA binding leads to a more extended protein-ligand complexes. The two compounds are expected to yield AS aggregates with different structural properties and toxicities, thus, reducing the amyloid fibrillogenesis.

2.5 AS CHARACTERISATION IN BIOPHYSICS

The structure of AS has been widely studied in biophysics, principally for its peculiar role in neurodegenerative diseases. Indeed, in the last years, the AS was considered the perfect model for further studies of IDPs conformations. Historically, the first techniques employed for AS structure investigations were common biophysical bulk assays sensitive to the presence of secondary structure [150]. More recently, very sensitive methods, such as native MS, single molecule techniques (both force spectroscopy and fluorescence-based) and computational approaches, i.e. MD, have been developed.

2.5.1 *Low-Resolution Biophysical Analyses of AS*

Early characterization of AS secondary structure utilized circular dichroism (CD) spectroscopy [151]. Analysis of the CD spectrum of free AS led to estimates of less than 2% α-helix structure and nearly 70% random coil content in solution. The founded Stokes radius was consistent with a globular protein of 57 kDa rather than 14.5 kDa expected for the AS. Small-angle X-ray scattering (SAXS) was employed to determine the radius of gyration produced by AS X-ray scattering curves of $\sim$ 41Å instead of 15.1Å estimated for a globular protein or 52Å calculated for a completely unfolded polymer of the same size [152, 153]. Thus, AS was found to lack significant secondary structure and to adopt a conformation which was slightly more compact than the extended random coil state. Fourier-transform infrared (FTIR) spectroscopy displayed a broad peak at 1650 cm^{-1}, supporting the notion of an high degree of structural disorder [153].

2.5.2 *NMR Spectroscopy-Based Analyses of AS*

Paramagnetic relaxation enhancement (PRE), which exploits the altered relaxation rates of nuclei and enables identification of interatomic interactions distances up to 25Å apart [150], was employed to detect the conformational structure of AS. The data, both experimental and simulated, obtained by PRE investigations, suggested non-random, long-range interactions between the N- and C-terminus of free AS, compatible with a partially compact structure [154, 155] or a continuum of dynamics and transient conformations [156].

2.5.3 *Native MS of AS*

Native MS has developed into a central tool for structural biology, conjugating its strong analytical power with structurally meaningful information. Indeed, intact proteins can be infused into the mass spectrometer and mild desolvation/ionization methods prevent

backbone fragmentation and can also preserve non-covalent interactions, such as those responsible for protein folding and binding. The analysis of charge states populated by globular and disordered proteins by native MS has shown effects of denaturants, stabilizers, metal binding, and protein:protein interactions [157]. The application of native MS to free AS protein in solution revealed multimodal charge-state distributions (CSDs), which are suggestive of a conformational ensemble populated by different conformers, in line with both the *in vitro* and *in vivo* evidence obtained for other studies [157, 158, 159, 160].

Native MS has described conformational responses of AS to alcohols, pH, and copper binding [122, 161] and also suggested that binding of DA and EGCG have distinct structural effects on AS soluble monomers [149, 162].

2.5.4 *Single Molecule Fluorescence of AS*

Fluorescence techniques are usually employed for studying the kinetics of aggregation of AS by turbidity or Thioflavin T (ThT) fluorescence essays. In particular, early stages of AS aggregation can be follow with fluorescence correlation spectroscopy (FCS), which is a technique that allows the determination of one molecule's size by analyzing its free diffusion [163]. However, these approaches do not permit the structural study of a single AS monomer in solution, due to their big signal fluctuations. Recently, single molecule fluorescence-based techniques (SMF) have started to have a key role in the conformational study of AS and other IDPs [164, 165, 166]. Among these, the single-molecule intermolecular Fluorescence Resonance Energy Transfer (smFRET) was employed to study the transition from monomers to small oligomers of both WT-AS and mutated AS [167] and WT-AS at different pH [168], providing unique insights in the description of initial aggregation.

2.5.5 *Single Molecule Force Spectroscopy of AS*

The co-existence of several species of the same protein in solution at equilibrium, some of which may be scarcely populated, makes analyzing IDPs a particularly challenging task using established biophysical techniques [169]. By contrast, SMFS techniques provide the unique ability to tackle this problem being able to reveal monomeric conformers, without suffering by averaging. SMFS have already been employed to monitor the activity and conformation of proteins with structural plasticity, resolving fast fluctuations (20-300 ns) at the monomeric level [170]. Moreover, the nanomanipulation at the single molecule level has the unique opportunity of directly measure the forces involved in stabilization of the intramolecular interactions,

from here called *mechanical stability*, and to directly correlate the mechanical stability of a IDP with its conformation.

In particular, AFM-based SMFS was widely employed to study the conformational equilibrium of monomers, and even to detect rare forms. The mechanical stability of a protein is strictly linked to its secondary structure, it is known that the β-strucure is usually more mechanically stable than α-helix and disordered conformations.

Several SMFS studies of AS employed the classical approach, in which the AS was placed in series with the marker I27 repeats [88, 89] or by inserting the AS as a guest inside a carrier protein of well-known mechanical properties (which usually displays an higher mechanostability with respect to the markers) [169, 171]. Both these approaches resolved different conformers for WT and mutant-AS, suggesting that the most mechanostable structure could be linked to the existence of a β-sheet structure while the weaker one was related to unfold and unstructured conformations. It is common to obtained other conformations, i.e. a mechanically weak intermediate state [88] or an higher mechanostable structure [169].

2.5.6 *Molecular Dynamics of AS*

In contrast to ordered proteins, IDPs are usually represented by a broader variety of structures in MD studies, reflecting the conformational freedom of the IDPs [172]. The majority of the MD studies focused on the early stages of AS aggregation, i.e. from monomers to small oligomers, or on the contribution of a specific component. Moreover, the results obtained from simulations of AS strongly depend on the set of force-field parameters employed to describe the energy of an IDP and its interactions with the aqueous solvent [173]. It is particularly common to found several different conformers, ranging from unstructured to more compact forms. For example, a recent work for AS free in solution described three different conformations: compact globular structures containing a few distinct secondary structure fragments (1) and two conformations characterized by reflecting long-range electrostatic interactions between N- and C-terminus (2) or intermediated sub-domains (3) [174].

3

DNA

3.1 BIOLOGY OF DNA

The DeoxyriboNucleic Acid (DNA) is a biomacromolecule composed by a two chains, which coil around each other to form an intermolecular double helix carrying all the genetic informations of an individual. The two DNA strands are linear polymer composed by *nucleotides*. Each nucleotide is composed of one of four nitrogen-contining nucleobases (purines: adenine [A] and guanine [G]; pyramidine: cytosine [C] and thymine [T]) and a desoxyribose-phosphate backbone (Fig.3.1). In the double-helix configurations, Adenine is always paired with Thymine via two hydrogen bonds and Guanine is always paired with Cytosine via three hydrogen bonds, while two consecutive nucleotides are connected to each other via covalent phosphodiester bonds between the sugar of one nucleotide and the phosphate of the next [64, 65].

3.2 STRUCTURE OF DNA

The discovery of the double-helix structure of DNA in 1953 by Watson and Crick [175], thanks to further X-ray diffraction images obtained by Rosalind Franklin, started the modern biology molecular. The Watson-Crick DNA structure is the so-called B-form, which is the fundamental and more widespread structure. Other DNA-structures, discovered years later, are A-form and Z-form.

Despite of the great variability of DNA sequence, which reflects the genetic information carried by each organism, the structure of DNA molecules follows the same base-pairing rule between the two strands. The nanomechanics of DNA is widely explored principally in biophysics. It is studied because of DNA peculiar characteristics: indeed the double-stranded DNA (dsDNA) biomolecule is stiffer than artificial polymers and than most of other natural polymers [176]. This rigidity comes from both the hydrogen bonds of base-pairing and the base-stacking, which is the main component in determining the bending stiffness.

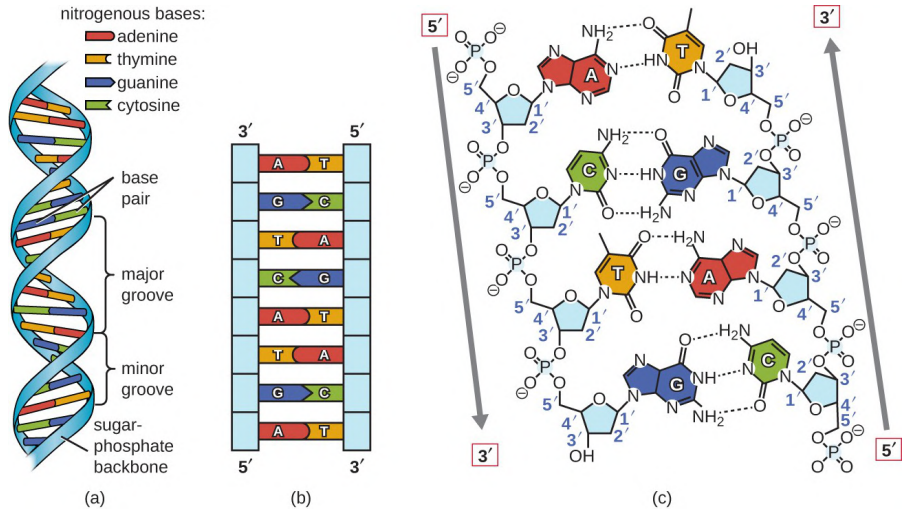

Figure 3.1.: (a) The sugar-phosphate backbones are on the outside of the double helix and purines and pyrimidines form the "rungs" of the DNA helix ladder. (b) The two DNA strands are antiparallel to each other. (c) The direction of each strand is identified by numbering the carbons (1 through 5) in each sugar molecule. The 5′ end is the one where carbon #5 is not bound to another nucleotide; the 3′ end is the one where carbon #3 is not bound to another nucleotide. Hydrogen bonds form between complementary nitrogenous bases on the interior of DNA (A-T and G-C couple). Fig. taken from Lumen Learning.

3.2.1 *B-DNA*

The canonical form of the DNA double helix in aqueous solution is B-DNA, a right-handed double helix where each base-pair (bp) has an angle of about 34° with the previous one [177, 178], with an helical period of is approximately 10.5 bp per turn and an helix rise of ~ 0.34 nm per bp. The external diameter of the hydrated double helix is about 2nm [179, 180, 181]. B-DNA double helix is not perfectly symmetrical, indeed it has two distinct grooves, a major groove 12aa wide and a minor groove 6aa wide. The geometry of the grooves has a very important biological role, because there are several proteins or drugs that recognize specific sites on DNA through the binding with these grooves [182]. The rigidity to B-DNA is quantified by a *persistence length* (Appendix D.3) of about 50nm, which is about fifty times larger than the one of artificial polymers, such as polystyrene [64].

3.3 PROTEIN:DNA INTERACTION

Almost all the functions carried out by the DNA molecules involve interactions with proteins, in specific (i.e. the protein could bind a specific DNA sequence) or non-specific way.

Among all the interactions existing between DNA and proteins, proteins involved into the transcription, replication and structural rearrangement and package of DNA are particularly important.

3.3.1 *From DNA to Chromosome*

The DNA in a human cell is approximately 2 m-long and is organized into a precise and compact structure, to fit into the nucleus of a cell of $\sim 5 - 10\mu m$ in diameter.

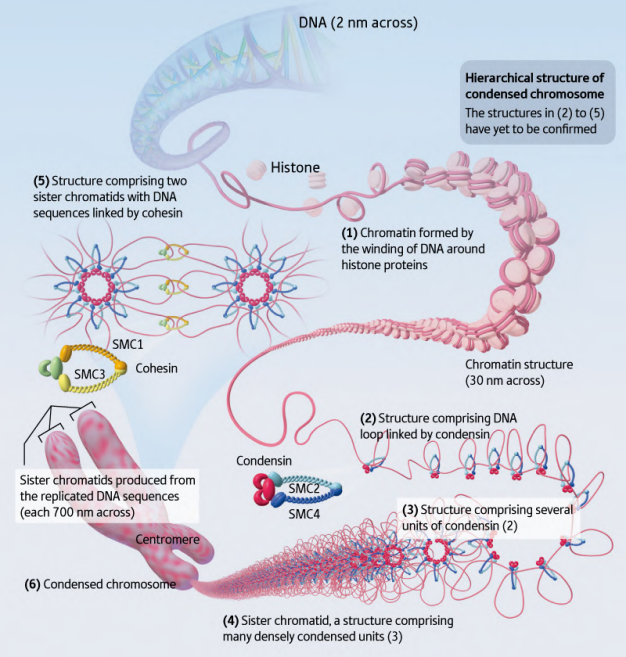

Figure 3.2.: Cartoon of structure arreangment of DNA into Chromosome. [Adapted from Riken©]

Consequently, DNA is combined with proteins and organized into a dense string-like fiber called *chromatin*. Each DNA strand wraps around groups of small protein molecules, the *histones*, forming a series of bead-like structures, the *nucleosomes*, connected by DNA strands (Fig.3.2). Due to this compaction, the negative charge of the DNA molecule is neutralized by the positive charge of the histone molecules,the inactive DNA can be folded into inaccessible locations until it is needed, but, on the other end, it must be able to decondense to perform its functions in gene expression and genome maintenance. Despite this historically picture of DNA condensation into chromosome, new insight suggest a condensin-mediate mechanism which can stochastic cross-link freely diffusing chromatin fibres [183]. The principal proteins that shape chromosomes into their definitive compact structure are the Structural Maintenance of Chromosomes (SMC) protein complexes [183].

3.3.2 *SMC Complexes*

The Structural Maintenance of Chromosomes (SMC) complexes, which include, among eukaryotes, *cohesin*, *condensin* and *Smc 5/6*, are the major components in chromosomes.

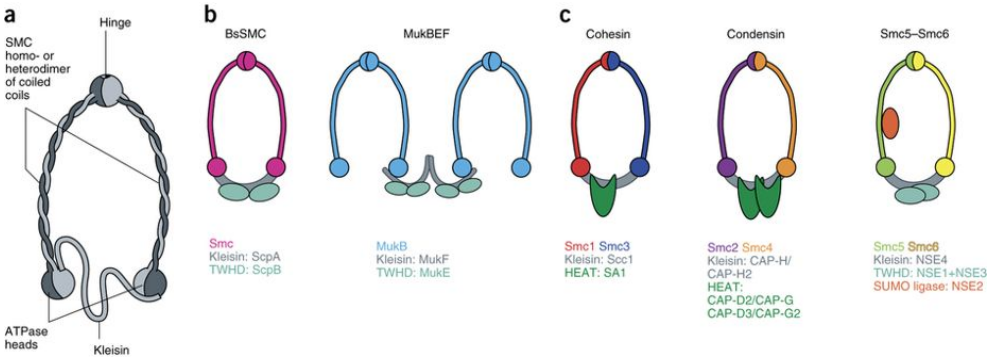

Figure 3.3.: (a) General architecture of SMC complexes (two SMC proteins connected at the hinge). In prokaryotes, the complex is a homodimer, whereas eukaryotic complexes are heterodimeric. The opposite ends of the SMC proteins possess ATPase activity. A kleisin subunit completes the ring. (b) Overview of prokaryotic SMC complexes. (c) Overview of eukaryotic SMC complexes. Fig. taken from [55].

These complexes have a ring-shape structure and it has been demonstrated that they topologically encircle DNA strands, by a ATP hydrolysis (Fig.3.3). The SMC complexes can trap more than one DNA strand, and they are indispensable in a plethora of different essential mechanism, as chromosome condensation, sister chromatid cohesion and DNA repair.

The first SMC complex was isolated by *Niki et al.* [184, 185] from an *E. Coli* mutant: the *mukB* protein, consisting in a amino-and-carboxyl-terminal globular ATPase domains separated by two long and flexible stretches of coil-coil.

Early on, the role of the SMC complexes in chromosome structural organisation and shape and segregation was proved, together with the understanding that in eukaryotes, they are the principal core of multi sub-unit protein complexes. Particularly, the cohesin complex is essential for mediate sister chromatid cohesion [186, 187], while the Smc5/6 plays a role in repair DNA [188] and the condensin is the main palyer in chromosome condensation [189, 190]. Although these three complexes are essential for life, they partially overlap their roles. It has been claimed that condensin is an ATP-powered molecular motor [54, 55], while no evidence have been observed for cohesin and Smc5/6 yet.

3.3.3 *Architecture of SMC complexes*

The SMC complexes are all characterized by a ring-shape structure, made up by two long stretches coil-coil of the two SMC subunits, connected to one end by a stable dimerisation interface, the *hinge*[55].

At the opposite end an ATP binding subunit is present (ATP-binding cassette, ABC). The heads are held together by a kleisin subunits and a ATP-dependent dimerisation [191, 192, 193, 194, 195].

This ring-shape led to the hypothesis that it can bind the DNA by a topological embrace, even if how the DNA enters and exits from the ring is not totally assessed.

3.3.4 *Condensin*

The structural maintenance of chromosome (SMC) protein complex condensin plays a central role in chromosome structure regulation [196, 197]. Like the other SMC complexes, the condensin shows a ring-like structure [198]. In particular, the condensin protein has an heterodimer of Smc2 and Smc4 subunits, connected at a globular *hinge* and fold back to form a 45-nm long flexible coiled coils *arms* and an ATPase domain at the other end. To complete the ring structure, the two Smc2 and Smc4 coils are connected by a kleisin-like protein (Brn1), which is lined to other two additional subunits, consisted by HEAT-repeat motifs (Ycg1 and Ycs4) (see Fig.3.4). As reporte in [198], the condensin protein could adopt several different shapes: while the the coils coils are always connected at the hinge, the other two ends of the Smc2 and Smc4 populate different conformational classes.

3.3.5 *Models of DNA:SMCs interaction*

All the SMC complexes are thought to promote genome architecture by physically linking distal chromosomal loci, but this bridging mechanism is still unknown [58]. In the last three years, several models have been proposed, trying to describe the interactions between SMC complexes and DNA.

Extrusion Models

The *loop extrusion model*, proposed by C. Dekker and coworkers [58, 54, 55], suggests that the condensin is actively creating the DNA loops. This model is supported by both dynamic polymers simulation and experimental evidences, and could explain the three-dimensional features of eukaryotic chromosomes. However, the loop extrusion model remains hypothetical, because the motor activity of proteins for driving loop extrusion is not yet identified. Recently, Hassler et colleagues [199] proposed three interplaying loop extrusions models

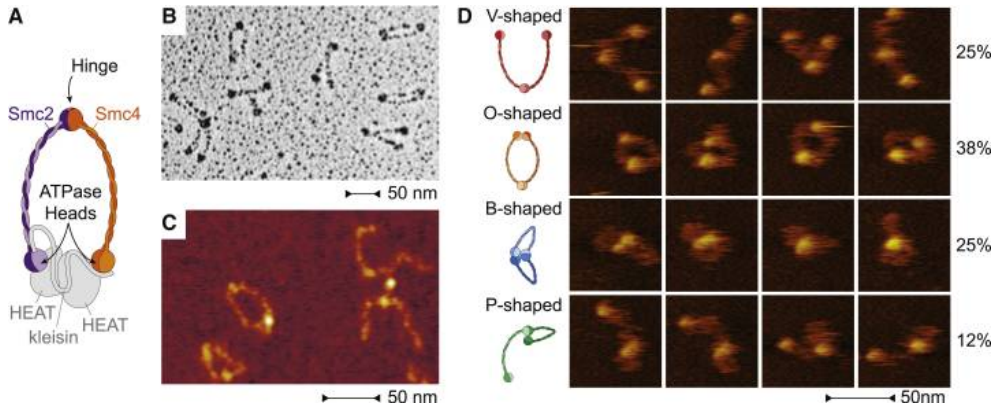

Figure 3.4.: A) Cartoon of the eukaryotic condensin complex. Smc2 and Smc4 heterodimerize via their hinge domains. The kleisin subunit associates with the Smc2 and Smc4 ATPase head domains to create a ring-like structure and recruits two additional subunits. (B) Example image of Smc2-Smc4 dimers imaged by rotary shadowing EM. (C) Example image of Smc2-Smc4 dimers imaged by dry AFM. (D) Example images of different conformational classes of Smc2-Smc4 dimers from high-speed liquid AFM movies. The frequency of each conformational class (as fraction of 1.795 total frames from 18 movies) is indicated. V-shaped, SMCs are connected at the hinge but the heads are not engaged; O-shaped, the heads are engaged with each other; B-shaped (butterfly), both heads are engaged with the hinge; P-shaped, one of the heads is engaged with the hinge. Fig. taken from [198].

based on ATP-hydrolysis-mediated changes in the geometry of the SMC coiled-coil arms, which can explain the ability to overcome the challenges of translocation on chromatin fibers, where nucleosomes and other DNA-bound proteins pose physically obstacles (Fig.3.5).

The Random Cross-Linking Model

The *random cross-linking model* claims that DNA loops might be stabilized by stochastic binding of two SMC complexes to bridge distal chromosome loci [200]. Moreover, the low rates of ATP hydrolysis with respect to other known nucleic acid motor proteins, implies that the SMC complexes would not be efficient in biologically time scales. However, this discrepancy can be explained if SMC proteins are able to take large steps, which is conceptually possible given their large size of > 50 nm.

The Sequential Walking Model

This model suggests that the two SMC ATPase head domains sequentially bind DNA (Fig.3.5 A) and, while one head attaches the complex stably to DNA, the other ATPase head contacts another binding site

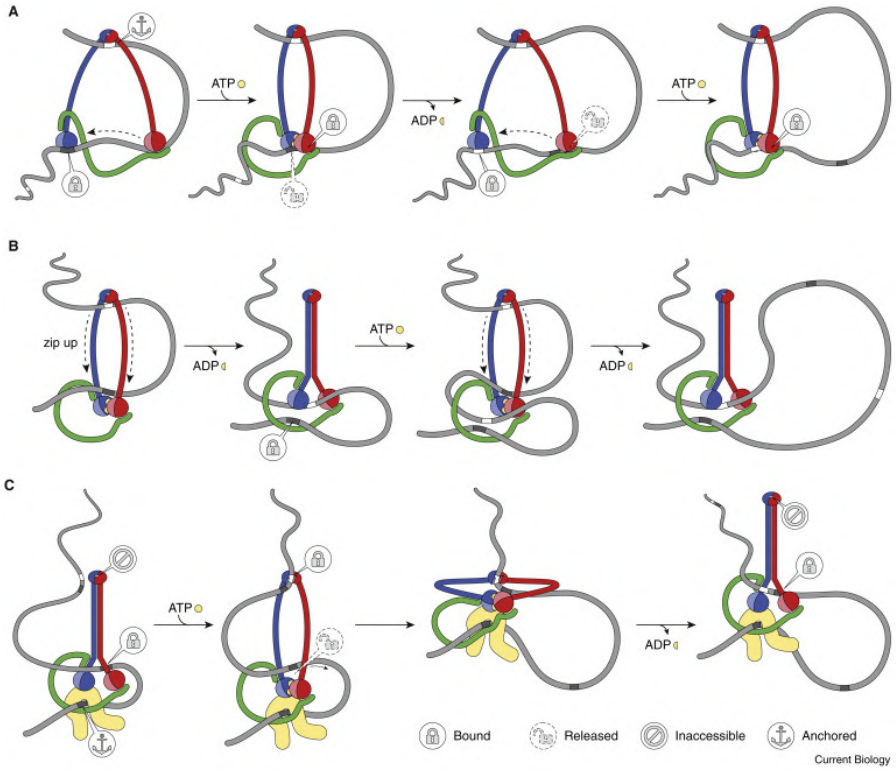

Figure 3.5.: (A) Sequential walking model. (B) DNA pumping model. (C) Extended scrunching model. Note that the sequential walking model makes no assumptions about the SMC coiled coil conformations, whereas the pumping model assumes stiff coiled coils that are under tension when bent open. The extended scrunching model postulates that SMC coiled coils alternate between stiff and relaxed states. Fig. taken from[199].

on the same DNA molecule. It is assumed that SMC–kleisin ring architecture remains intact until the DNA binding of the second ATPase head triggers a conformational change that releases the first head position to start the next step of the translocation cycle. This model is completely compatible for the 16-nm step sizes that have been calculated for condensin's movement along DNA [58] while bigger steps could be explained by a relaxed DNA, easier to compact. This model fails in describing the translocation of the condensin in the same direction for long distances and an alternate binding of the complex to DNA is not clear in biochemical assays.

The DNA Pumping Model

A second model relies on pseudo-topological DNA loop trapping between the two open SMC coiled-coils (Fig.3.5 B). After an ATP hydrolysis, the coiled coils zip up and a portion of the entrapped DNA loop is pushed towards the free ATPase heads. Later on, the heads re-engage, the SMC coiled coils return to the open conformation ring

and the base of the DNA loop is now entrapped within a second subcompartment between the engaged ATPase head domains and the kleisin subunit. This non-conventional model fails in describing how the DNA loops could be caught between the SMC coiled coils and how the mono directionality is ensured.

The Extended Scrunching Model

The scrunching model (Fig.3.5 C) was initially proposed by Dekker and coworkers [58] and extended later on by Hassler et al [199]. In this model, the ATPase cycle powers the conformational transitions between coiled-coil opening and closure providing DNA access to either of the two DNA binding sites. Stiffening of the SMC coiled-coil arms provides the power stroke that places the hinge at a ~ 50-nm distance to the head domains. In this model, loop extrusion would be achieved through DNA attachment to an additional, constitutive anchor site, which ensures that condensin drags the DNA helix into a loop while it translocates.

This model explains some evidences of DNA loop extrusion, and it might be able to explain condensin steps significantly larger than 50 nm, and it is not dependent to preformed DNA loops, but instead allows SMC complexes to initiate loop formation. Therefore, this model does not explain how condensin backwards steps could be prevented and how would the hinge and head domains know that they have come into proximity before handing over DNA binding.

Elbows models

More recently, Bürmann and coworkers [201], showed that two SMC complexes, MukBEF and cohesin, present a dynamic coiled-coil discontinuity, called the *elbow*, near the middle of their arms that permits a folded conformation. They proposed a new model which implies the bending at the elbow brings closer the hinge dimerization domain and the head–kleisin module, leading on a model that include a large conformational change in SMC arms, and a relative movement between DNA contact sites during DNA loading and translocation. The two proposed models for DNA translocation and loop extrusion involve an intermediate folded state (Fig.3.6). The first model, *inchworm translocation* uses distance changes between two DNA binding sites, either of which might be a topological entrapment device or ring, formed by the SMC heads and the kleisin. Folding at the elbow might cause the distance change. The second model, the *segment-capture* mode, describes the translocation using the segment-capture mechanism that enlarges a loop held in a bottom chamber by merging with a smaller loop captured in a top chamber. Folding at the elbow might drive DNA from top to bottom.

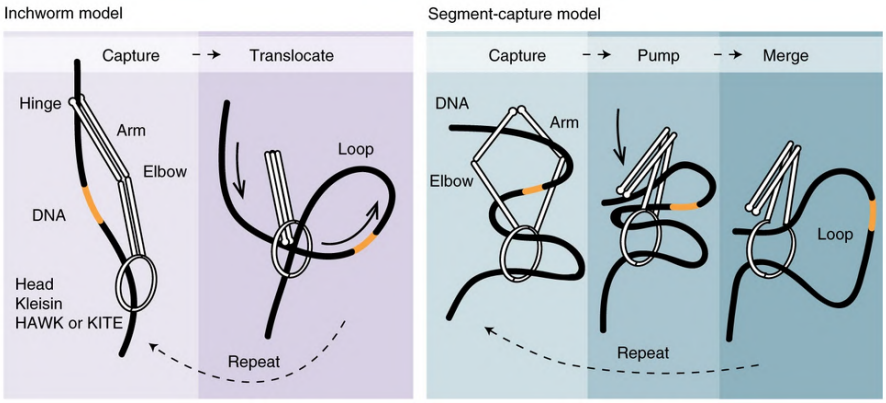

Figure 3.6.: The two elbow-based modes: the inchworm mode (Right) and the Segment-capture model (Left). Fig. adapted from[201].

Despite the great number of models proposed, the real mechanism of SMC trapping DNA is not fully explained right now. Some evidence in loop extrusion models have been reported for condensin and yeast cohesin, while recently Gutierrez-Escribano and colleagues [202] have reported a DNA bridging activity that requires ATP and is conserved from yeast to human cohesin, by two distinct classes of bridges at physiological conditions, a *permanent bridge* able to resists high force (over 80 pN) and a *reversible bridge* that breaks at lower forces (5-40 pN). The models which can explained this feature are a *two-compartment* model, the *3-bar-buckle and multiple compartment (pretzel)* models, or different compartments of two cohesin complexes, *the handcuff model*. However, since these last three models were proposed for the explaining the mechanism of cohesin protein, they won't be exhaustively explained here. Indeed, even if the finding of a unique SMC:DNA interactions model is still a challenge, it is also possible that two different SMC complexes have different ways to compact the DNA, explaining also the different function they have into the living organism.

Part II

DEPICTING CONFORMATIONAL ENSEMBLES OF α-SYNUCLEIN BY SINGLE MOLECULE FORCE SPECTROSCOPY

4

α-SYNUCLEIN SINGLE MOLECULE FORCE SPECTROSCOPY

Intrinsically disordered proteins (IDPs) play crucial regulatory roles in biological systems and lack a well define and specific tertiary structure under physiological conditions *in vitro*, showing an elastic and extended conformation (Section 2.3)[203, 204, 205, 206]. The molecular characterization of IDPs requires description of the heterogeneous conformational ensembles populated by the disordered polymers in solution. Trying to depict such a complex frame by using traditional structural technique, like x-ray crystallography, is still really challenging, suffering from the limitations of average parameter assessment, intrinsic to bulk methods On the other ends, the relatively recent single-molecule approaches could offer information on dynamic and heterogeneous ensembles, capturing also the distinct and less populated states which are usually hidden by the most abundant conformations [207, 208, 209, 210].

In the case of the human amyloidogenic IDP α-synuclen (AS), at least three major conformational states can be recognized by using a SMFS approach [88, 89, 169]: Random Coil (RC), collapsed states stabilized by Weak Interactions (WI), and compact conformations stabilized by Strong Interactions (SI).

Pure AS in vitro suffers, like the majority of IDPs, from loss of structure: in the absence of interactors, it is largely unstructured at neutral pH, presenting only a small fraction in collapsed states of different compactness, as revealed by NMR spectroscopy [211] and small angle X-ray scattering [212]. DA and EGCG are known to bind AS and redirect the aggregation pathway towards soluble oligomers with different structure and toxicity [213, 214]. Native MS to free AS revealed multimodal charge-state distributions (CSDs), which are suggestive of a conformational ensemble populated by different conformers [160], also suggesting that binding of epigallocatechin-3-gallate (EGCG) and dopamine (DA) have distinct structural effects on AS soluble monomers [149, 162]. DA preferentially binds and stabilizes an intermediate form of the AS while EGCG promotes the population of the most compact AS conformer. Both the ligands have a similar fibrillation-inhibition effect, but it has been shown that they forms oligomers of different toxicity and structure.

The aim of this work is to describe AS conformational ensemble and the effects of ligands by a SMFS technique, to test the effect of ligand binding in solution and help interpretation of the AS and IDPs data in the complicate frame of the heterogeneity population of structural forms.

For this reason, in this part of the Thesis I used an AFM-based single molecule force spectroscopy approach to depict the conformational states of an intrinsically disordered protein, the AS. Since some ligands, such as EGCG [213] and DA [214], are known to affect the fibrillation process of AS, I used this single molecule technique to investigate the effect of EGCG and DA on the conformational ensemble of the WT AS.

Moreover, it has been revealed from several studies that the presence of point mutations, linked to familial Parkinson's disease (PD), correlate with the gaining of structure and therefore with AS aggregation. Furthermore, SMFS approaces had revealed the presence of distinct conformers of the molecular ensemble and structural effects of point mutations linked to familial PD [206, 89]. Therefore, I performed single-molecule force spectroscopy (SMFS) studies also on three of this AS point mutants (A30P, A53T and E83A).

Data related to AS:ligand interactions presented here have been published in [1].

5

MATERIALS AND METHODS: CONFORMATIONAL ENSEMBLES OF AS

Briefly, the final construct is a $I27_4 - AS - I27_4$ polyprotein (with AS = WT, A30P, A53T and E83A). At the C-term two modified Cys form a thiol (covalent) bound with the gold substrate, while at the N-term 6x His-tag for protein purification. The polyprotein preparation could be divided into two main part, each of them fully explained in the next sections:

- Molecular cloning to obtain a final expression vector encoding the sequences for an AS protein flanked by flanked by four repetitions of titin immunoglobulin-like domain (I27) at the N-terminus and at the C-terminus.
- Protein expression in *E. Coli* strands and purification

5.1 MOLECULAR CLONING

Briefly, in order to obtain the $I27_4 - AS - I27_4$ polyprotein (with AS = WT, A30P, A53T and E83A) we took advantage of the pRSet.A $(I27)_8$ vector already successfully used to study the unfolding of a I27 octamer [215]. The final construct consist of a single AS molecule, flanked by four repetitions of titin immunoglobulin-like domain (I27) at the N-terminus and at the C-terminus. The cDNA of the human AS (NP-000336) was cloned in the pRSet.A$(I27)_8$ expression vector, taking advantage of the NheI restriction site placed in the middle of $(I27)_8$ encoding sequence (for more details see Fig.C.2 in Appendix C). A mutagenic PCR was performed on the pEGFP-AS vector to delete the start and stop codons and to insert an NheI restriction site at both extremities of the AS gene. The PCR was carried out using the Q5 ®High-Fidelity DNA Polymerase (NEB, cat. # M0491).

5.1.1 *DNA Electrophoresis on agarose gel*

To confirm the success of each step (mutagenic PCR [Fig.5.1], colony PCR [Fig.5.2], ligation, digestion *etc*), the DNA electrophoresis on agarose gel was widely used. The 1% agarose gel was prepared by

using 1 g of agarose (Euroclone, Gelly Phor, Cat.no. EMR010500) dissolved in 100 mL of Buffer TAE 1x [40 mM Tris Base, 1mM Na_2 EDTA, 20 mM acetic acid and water to volume, pH 8.6] and 0.5 mg/µL ethidium bromide (EtBr). The DNA samples were prepared for loading by adding Gel Loading Dye, Purple 6X [(NEB, Cat. no. B7024S) 2.5% Ficoll ®-400, 10 mM EDTA, 3.3 mM Tris-HCl, 0.08% SDS, 0.02% Dye 1, 0.001% Dye 2: pH 8]. The electrophoretic run was conducted at 100 V in TAE buffer 1x. The bands were visualized by exposing the gel to ultraviolet light (λ = 312 nm for EtBr) by means of Amersham Imager 600 (GE Healthcare, Life Sciences).

5.1.2 *Polymerase chain reaction (PCR)*

Mutagenic PCR of AS cDNA for WT and mutants

The pEGFP plasmid containing the cDNA sequence of SNCA WT (kindly provided by Professor Cappelletti, UniMi) and the pET11a plasmid containing the cDNA sequences of SNCA mutants (kindly provided by Professor Legname, SISSA, Trieste) were used as template for a mutagenic *Polymerase chain reaction* (PCR) to obtain an exponential amplification of cDNA containing the desired restriction site at both the ends of the sequence.

The primers that were used for PCR mutagenesis are:

Forward:5′ AAAA**GCTAGC**GATGTATTCATGAAAGGAC 3′
Reverse:5′ AATT**GCTAGC**GGCTTCAGGTTCGTAG 3′

The bases in bold underline the restriction site for the enzyme NheI. Primers were designed also with the aim to delete the start codon, ATG, and the stop codon, TAA, present in the original sequences, via mutagenic PCR. The PCR was peformed in a Mastercycler nexus (Eppendorf) thermocycler in a final volume of 25 µL. Final PCR mixture was: 1 ng of DNA template, 0.5 µM primers, 0.2 mM dNTPs and 0.6 units of Q5 DNA polymerase in Q5 reaction buffer, at a final concentration of 1x. The PCR cycles and the temperature and time of denaturation, annealing and elongation are resumed in the Tab.5.1. The fragments of SNCA WT and mutants, amplified by PCR mutagenesis, were purified by using QIAquick PCR Purification kit (Qiagen, Cat. No. 28106).

In Fig.5.1, the ~420 bp inserted SNCA (WT and mutant) amplicon band was clearly placed just below the band of the marker that corresponds to 500 bp. Moreover, a unique intense band indicated that the amplifing PCR was abundant and highly specific.

Fragments of WT and mutagenic SNCA were then cloned in the pcDNA3.1 plasmid to permit the sequencing process. As a matter of fact, also in the presence of high fidelity DNA polymerase the PCR process may include undesired mutations. It was impossible to carry out this sequencing process in the final vector pRSet.A($I27_4 - AS -$

	Step	Temperature (°C)	Time (s)	Number of cycles
Step 1	Initial denaturation	98	30	1
Step 2	Denaturation	98	10	10
	Annealing	55	20	
	Extension	72	30	
Step 3	Denaturation	98	10	25
	Annealing	68	20	
	Extension	72	30	
Step 4	Final Extension	72	120	1

Table 5.1.: Condition of mutagenic PCR

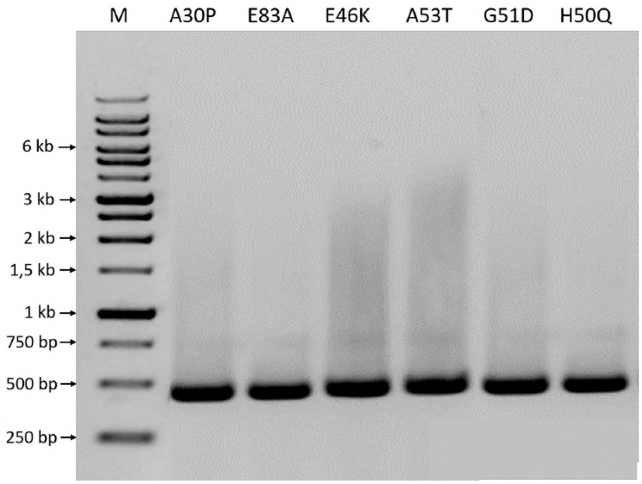

Figure 5.1.: PCR mutagenesis of SNCA mutants. 1% agarose gel, stained with EtBr. M(marker)= 1kb DNA ladder (0.25 µg). The base pairs relative to every band are reported on the left. On the top of each lane (from 2 to 7)it is indicated the name of every mutant used as template for the amplification. For each sample 25 µL of PCR mixture were loaded.

$I27_4$) because the starter point of the chosen primer would recognize a part of the I27 sequence, which was repeated 8 times.

5.1.3 *Digestion and ligation*

In order to insert the sequence encoding for the AS inside the pcDNA3.1, both the amplicons and the plasmidic vector have been digested with the NheI enzyme. Briefly, 2 µg of DNA fragments of SNCA WT and mutants, and 3 µg of the empty vectors (pcDNA 3.1 or pRSet.A $I27_8$) were subjected to digestion using 10 units of the restriction enzyme *NheI* [New England Biolabs (NEB) R0131S, 10.000 U/mL] in Cutsmart buffer 1x (50 mM Potassium Acetate, 20 mM

Tris-acetate, 10 mM Magnesium Acetate, 100 µg/ml BSA pH 7.9) for 16 hours at 37°C. Then, the enzyme was inactivated by heating the mixture at 65°C for 20 minutes. In the case of the plasmidic vectors, these were dephosphrylated by adding 10 units of CIP (*Alkaline Phosphatase Calf Intestinal*, NEB, M0290S, 10^4 U/mL) to the previously enzime-inactivated mixture, at 37°C for 20 minutes. Finally, the vectors were collected by using the PCR purification kit. The ligation was normally performed using ~100 ng of plasmid and the insert in a molar ration of 1 to 3, 120 units of T4 DNA Ligase (NEB, M0202S, 400 U/µL) in T4 DNA Ligase Buffer 1x (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT pH 7.5) and by adding ATP (Adenosine 5'-triphosphate disodium salt hydrate, Sigma, MW = 551.14 g/mol) at 1 mM final concentration. The ligation was incubated at 16°C over night.

5.1.4 *Preparation of competent cells*

The The next step is the transformation of the DH5α *E. Coli* strand with the ligation product. The different bacterial strains of *Escherichia coli* (*E. Coli*) that were used in our experiments, DH5α, were kindly provided by Professor De Michelis, UniMi. These cells were turned into competent cells through calcium chloride ($CaCl_2$ protocol, which is a rapid method that gives a higher yield of competence. Firstly, an overnight culture of liquid LB (10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl), a commonly used nutrient-rich media was inoculated at 37°C. Then, in a ratio of 1:10 this pre-culture were inoculated in 25 mL of liquid LB and grown until the culture reached an optical density (OD_{600}) ~0.4. The culture was then cooled on ice for 5 minutes and centrifuge at 4°C at about 2500 g for 10 minutes (by using the Eppendorf centrifuge 5804R). Afterwards, the supernatant was removed and the pellet was resuspended in a cold and sterile solution of Buffer 1 (100 mM $CaCl_2$, 70 mM $MgCl_2$, 40 mM sodium acetate pH 5.5), half of the volume of the beginning culture, and kept on ice for 45 minutes. Finally, the second pellet (obtained after a second centrifugation) was resuspended in a cold and sterile solution of Buffer 2 (100 mM $CaCl_2$, 70 mM $mgCl_2$, 40 mM sodium acetate pH 5.5, 15% glycerol), 1/20 of the volume of the beginning culture. The competent cells were then ready to be used or stored at −80°C.

Transformation by heat shock

50 µL of competent cells were incubated either with the ligation products or about 1-10 ng of plasmidic vector at 0°C for 20 minutes and then left at 42°C for 1 minute. Immediately after, the cells were inoculated at 37°C for one hour in 1 mL of liquid LB and were then plated in solid LB agar containing the appropriate antibiotic for the plasmid selection. The plates were then incubated at 37°C over night. Af-

	Step	Temperature (°C)	Time	Number of cycles
Step 1	Initial denaturation	95	5 min	1
Step 2	Denaturation	94	1 min	10
	Annealing	50	20 s	
	Extension	68	1 min	
Step 3	Denaturation	94	20 s	20
	Annealing	55	20 s	
	Extension	68	1 min	
Step 4	Final Extension	68	5 min	1

Table 5.2.: Condition of Colony PCR

ter the transformation, the cells were plated on LB-agarose gel with in the presence Amp antibody. The pcDNA3.1 vector contains the AMP-resistance sequence, so only cells which acquired the pcDNA3.1 should grown in this medium.

Colony PCR

For the colony PCR, each positive colony, coming from a ligation process, was dissolved in 20 µL of sterile water, and 10 µL of this mixture were used as a template for the PCR. The primers used for the colony PCR were the same as the one described in Sec.5.1.2 while the polymerase was the Taq-DNA polymerase (Euroclone, EURO TAQ - Taq DNA Polymerase, EME010001, 5000 U/mL) in its reaction buffer (200 mM Tris-HCl pH 8.4, 500 mM KCl). The colony PCR was peformed in a Mastercycler nexus (Eppendorf) thermocycler in a final volume of 25 µL. Along with the colony mixture, the PCR mixture was made of 0.5 µM primers, 0.2 mM dNTPs, 5 mM $MgCl_2$ and 1.25 units of Taq DNA polymerase in the EUROTaq Buffer 1x. The PCR cycles and the temperature and time of denaturation, annealing and elongation are resumed in the Tab.5.2. The samples were then stored at 4°C until their next use. To verify the colony PCR, electrophoresis on agarose gel was performed as described in Sec.5.1.1m see Fig.5.2.

5.1.5 *DNA extraction*

Miniprep DNA preparation

The plasmidic DNA was then amplified in the E. coli colonies and picked up to further characterisation. In the case of Miniprep DNA preparation, the plasmidic DNA extraction was performed by using the QIAprep Spin Miniprep Kit (Qiagen, Cat. no. 27104). One positive colony was incubated in 3-6 mL of liquid LB with the appropriate antibiotic and grown at 37°C over night. The cells were collected by

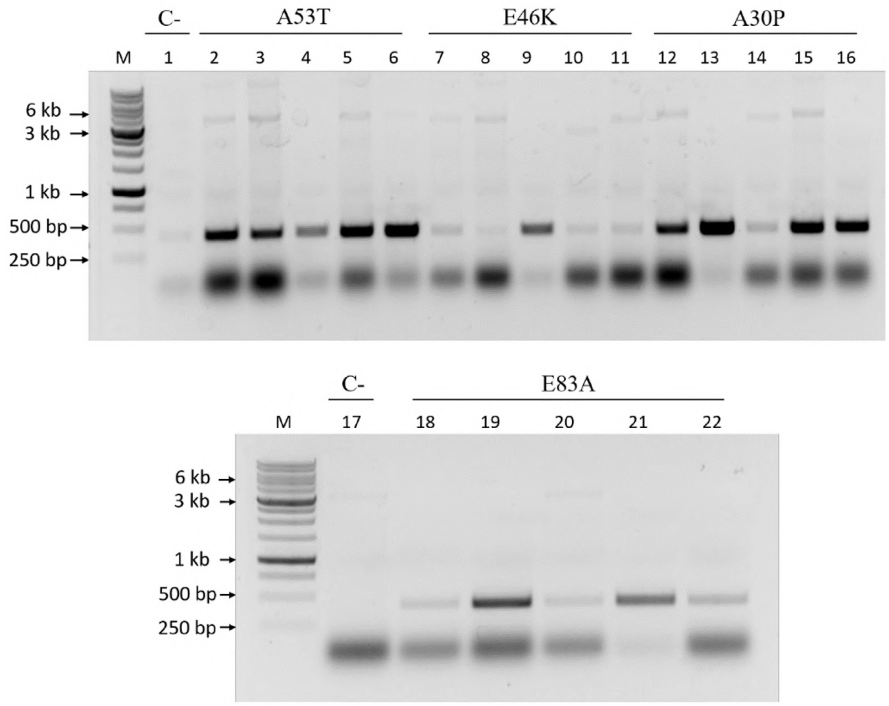

Figure 5.2.: **Colony PCR for the screening of some SNCA mutants in pcDNA 3.1.** 1% agarose gel, stained with ethidium bromide. M = 1kb DNA ladder (0.25 µg); the base pairs relative to every band was reported on the left. On the top of each lane (2-7) the name of every mutant used as template for the amplification was reported. For each sample 25 µL of PCR mixture were loaded.

centrifugation of the culture at room temperature at 2348 g for 3 minutes. The pellet was then treated as indicated from the manufacturer and the DNA was eluted in 40-50 µL of Elution Buffer (10 mM Tris-Cl, pH 8.5). The concentration of the DNA was then measured by Nanodrop One Microvolume UV-Vis Spectrophotometer. The final samples were then sequenced to control that no undesirable point mutations occurred. Sequence were aligned with *Clustal Omega* software and *Expasy* website.

Ethanol precipitation of DNA

A second purification protocol via *Ethanol precipitation* was also used as an alternative to section 5.1.5. The pellet of the cells was resuspended with 250 µL of Resuspension Buffer (50 mM Tris-Cl, 10 mM EDTA, 100 µg/mL RNase A, pH 8.0). Then, 250 µL of Lysis Buffer (200 mM NaOH, 1% SDS) was added to the mixture taking care that the reaction would not last more than 5 minutes, in order to avoid aspecific DNA disruption. Afterwards, 350 µL of Neutralization Buffer (4.2 M Gu-HCl, 0.9 M potassium acetate, pH 4.8) were added to stop

the lysis process. The whole mixture was centrifuged at about 16000 g at 4°C for 10 minutes. The supernatant was collected. Pure ethanol, 2.5 volumes of the supernatant, was added to the samples, which were incubated at −20°C for 30 minutes. Later, the samples were centrifuged at about 16000 g at 4°C for 30 minutes. The supernatant was removed and the pellet, containing the plasmidic DNA, was washed two times with 500 µL of ethanol 70%. The pellet was then dried over night. Hereafter, the precipitated DNA was resuspended with 50 µL of Elution buffer (10 mM Tris-Cl, pH 8.5) and stored at -20°C.

5.1.6 *Gel elution*

Positive samples were then digested by NheI and purified by means of gel elution to be sub-cloned on the final polyprotein plasmid (pRsetA). For the gel elution, the DNA fragments of interest, once run on 1 % agarose gel, were cut away taking advantage of an UV transilluminator (UVP Transilluminator PLUS, Analytikjena). For the next steps the QIAquick Gel Extraction Kit (Qiagen, Cat. no. 28704) was used. Assuming that 100 mg of gel equals 100 µL , for one volume of gel three volumes of Buffer QG (Solubilization Buffer with unknown composition) were added to the gel and incubated at 50°C until the gel slice was completely melted. Then, after adding one gel volume of pure isopropanol to the solution, it was applied to the QIAquick spin column and spinned at about 16000 g for 1 minute. The flow through was removed and 750 µL of Buffer PE (Wash Buffer: 10 mM Tris-HCl, 80% ethanol, pH 7.5) were added. Then the colomn was spinned and the flow through was removed again. Finally, the DNA was eluted from the column by adding 40-50 µL of Elution Buffer (10 mM Tris-Cl, pH 8.5) and stored at -20°C. Then, we performed the sub-cloning process of the SNCA amplicon (for bot WT and mutant) in the pRsetA vector, encoding for the polyprotein. The plasmid was digested with NheI and ligated with the SNCA amplicon. Positive Colonies were then selected by a new colony PCR.

5.1.7 *Test digestion*

To control the correct insertion of the SNCA amplicon in the final vector, also considering the 5'-3' direction, several tests were performed. The DNA extracted from the positive colonies thus was subjected to two different digestion tests: one single digestion with *KpnI* restriction enzyme and a double digestion with *KpnI* and *NdeI* restriction enzymes. 200-300 ng of plasmidic DNA (see Results chapter [METTERE QUI RIF] for more details) were subjected to digestion using 10 units of the restriction enzymes *NheI* (NEB, R0131S, 10.000 U/mL), *KpnI* (NEB, R3142S, 20.000 U/mL), *NdeI* (NEB, R0111S, 20.000 U/mL) and/or *BamHI* (NEB, R3136S, 20.000 U/mL) in Cutsmart buffer 1x for

1-2 hours at 37°C. The enzymes were inactivated by adding the Gel Loading Dye (see section 5.1.1) to the samples.

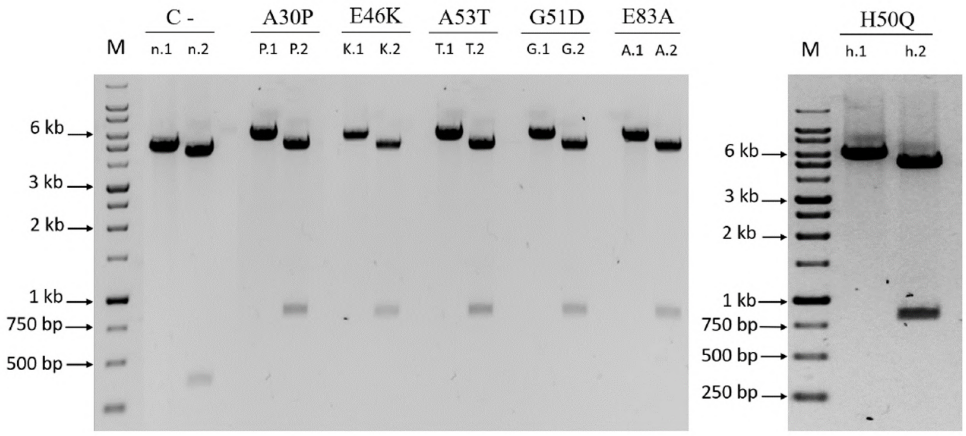

Figure 5.3.: Digestion tests for the screening of AS mutants in pcDNA 3.1. For each sample 200-250 ng of digested DNA were loaded on 1% agarose gel, stained with ethidium bromide. M = 1kb DNA ladder (0.25 µg); the base pairs relative to every band are reported on the left. C- = digestion test performed on DNA extracted from a colony grown on the plate relative to the negative control. On the top of each lane is reported the name of every mutant subjected to digestion tests: .1 = digestion with *KpnI*, .2 = double digestion with *KpnI* and *NdeI*.

The single digestion with *KpnI* should have got a linearized DNA fragment of ~ 6000 bp in the case of the positive clones and one at about ~ 5500 bp in the case of the empty vector (Fig.5.3, indicated as X.1 line). Instead, for the double digestion, we expect in both cases a band around 5000 bp relative to the vector backbone and a fragment of ~ 860 bp in the case of the positive clones, while a band of 437 bp in the case of negative control [Fig.5.3, indicated as X.2 line]. As shown in Fig. 5.3 all the tested colonies carried the pcDNA3.1 vector containing the cDNA codifying for the AS mutant fragment.

5.1.8 *Screening of the AS direction of insertion in pRSet.A* $I27_8$ *vector*

The last step of the molecular cloning was the control of the direction of insertion of the SNCA fragment in the pRSet.A $I27_8$ vector. Indeed, since the fragment of AS cDNA is flanked by the restriction site for *NheI* at both ends, during the cloning process it could have been inserted in the correct $5' - 3'$ direction or in reverse complementary direction.

For this reason, we performed a test digestion using the restriction enzyme *BamHI*, which has a single restriction site within the cDNA sequence of SNCA and one in the pRSet.A $I27_8$ vector (see map C.2). In the case of the correct insertion direction, the digestion test by

BamHI enzyme should give a band of 1472 bp, while in the other case a band of 1180 bp.

The digested samples were loaded on 1% agarose gel for the electrophoretic essay [Fig. 5.4 and C.4 in Appendix C. For all of the mutants at least one clone presented a digestion profile comparable to the positive control: four for the A53T mutant (# 1, # 3, # 6 and # 7) and three for the A30P mutant # 2, # 3 and # 4) [Fig. 5.4].

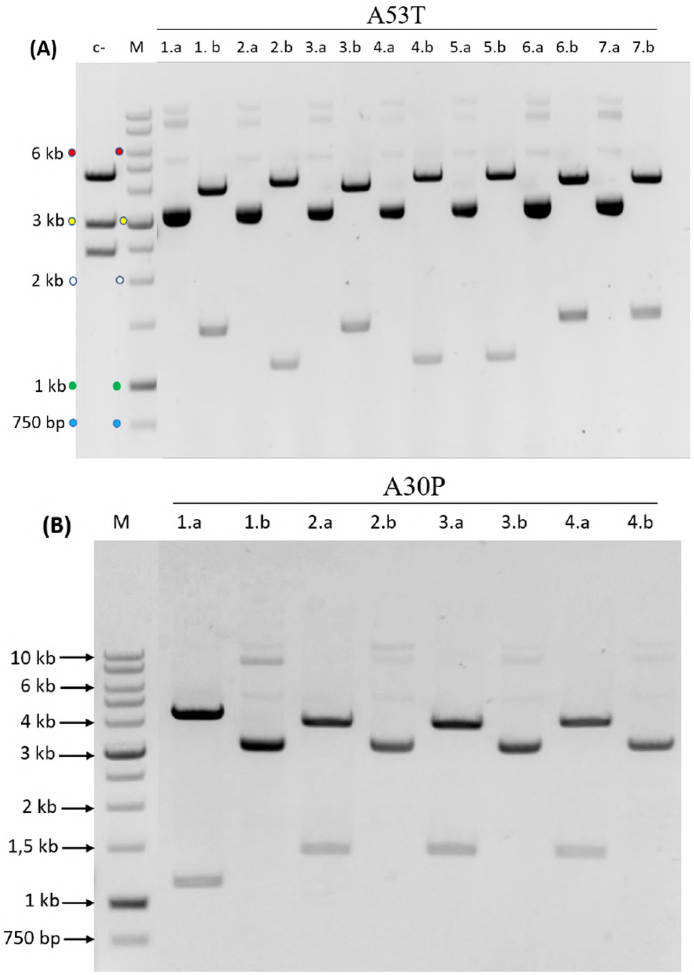

Figure 5.4.: **Test digestion for the screening of the AS mutant direction of insertion in pRSet.A** $I27_8$ vector. For each sample 200-250 ng of digested DNA were loaded on 1% agarose gel, stained with ethidium bromide. M = 1kb DNA ladder (0.25 µg); the base pairs relative to every band are reported on the left. C- = digestion test performed on DNA extracted from a colony grown on the plate relative to the negative control. C+ = digestion test performed on pRSet.A $(I27)_4$-SNCA WT-$(I27)_4$. For all the clones also the non digested DNA was loaded on the gel: .a = non digested, .b = digested with *BamHI*. A: A53T; B: A30P.

After sequencing, the pRSet.A $I27_4 - AS - I27_4$ vector was used to transform BL21(DE3) *E. coli* cells. Transformed cells were grown in Luria-Bertani medium at 37 ^{O}C until they reached an OD 600 of 0.4-0.6 and the expression of the polyprotein was induced overnight at 22 ^{O}C by the addition of 1 mM IPTG. Cells were subsequently harvested by centrifugation and resuspended in lysis buffer (50 mM Na_2HPO_4, 300 mM NaCl, 10 mM imidazole, 4% Triton®X-100, and 0.5 mM phenylmethylsulfonyl fluoride) before sonication on ice. The purification was performed by gravity flow column ion metal affinity chromatography (IMAC), taking advantage of the 6xHis-tag present at the N-terminus of the polyprotein. The soluble fraction of cell lysate was incubated on Ni+ - NTA resin (Roche, cat. # 05893682001) for 1 hour at 4 ^{O}C with gentle agitation. The washing step was carried out in 50 mM Na_2HPO_4, 300 mM NaCl added with 20 mM imidazole, elution was achieved in the same buffer, added with 250 mM imidazole. The presence of the protein in the eluted fractions was verified by SDS-PAGE on a 4-12 % polyacrylamide gel (Invitrogen®ThermoFisher Scientific, cat. # NW04120BOX) stained with Coomassie Brilliant Blue.

5.2.1 *Protein expression and purification*

The pRSetA $I27_8$ plasmidic vector, containing the cDNA codifying for AS, WT and mutants, was used to transform BL21 (DE3) *E. Coli* strain for protein expression. The BL21 (DE3) strain has been widely used to express recombinant protein. It carries a DE3 recombinant phage harboring the T7 RNA polymerase gene, regulated by the Lac operon. The T7 RNA polimerasi, upon IPTG induction, can direct high-level expression of the genes under the control of the T7 promoter [216]. In the absence of lactose, the lac repressor protein (LacI) binds to the operator sequence in the Lac operon, avoiding the transcription of T7 RNA polymerase and, therefore, all the genes under its regulation. However, in presence of lactose, it binds to LacI, releasing the repression of T7 RNA polymerase leading to the expression of all the genes that are under control of the T7 promoter. We used IPTG, a structural mimic of lactose able to bind to the lac repressor to induce gene expression. Unlike lactose, IPTG is not part of any metabolic pathways, meaning that it is not used by the cell. This peculiarity ensures that the concentration of IPTG added remains constant, making it a more useful inducer of the lac operon than lactose itself [217].

One of the positive colonies was incubated in 10 mL of liquid LB with ampicillin 100 µM and the culture was grown at 37°C over night. This pre-culture was then inoculated with a ratio of 1:20 in 200 mL of liquid LB with ampicillin and IPTG (Isopropyl β-D-1 thiogalactopy-

ranoside) at a final concentration of 1 mM and grown at 22°C over night to slow down the cell metabolism. Hereafter, the cells were collected by 1500 g centrifugation at 4°C for 10 minutes. The pellet containing the cells was freezed at −80°C at least for 2 hours.

For the protein purification, the iced pellet of BL21 was resuspended in 8-10 mL of lysis buffer (10mM imidazole, 50mM sodium phosphate dibasic and 300mM NaCl). Then, 0.5mM PMSF (Phenylmethanesulfonyl fluoride) and Triton X-100 4% were added to the resuspended cells.

The cells were further lysed by sonication by using SONICS Vibra cell ultrasonic processor. Afterwards, the soluble fraction has been recovered by centrifugation at 6000 g at 4°C for 10 minutes.

The purification of the protein of interest was done by IMAC (Immobilized Metal Affinity Chromatography). As the proteins contain an His-tag at the N-terminal, a Ni^{2+}-NTA resin (Roche, complete His-Tag Purification Resin) was used. 1mL of the resin was washed three times in Lysis Buffer for its equilibration. To have a higher range of purification, the soluble fraction was mixed with the resin and incubated at 4°C for one hour.

Aftewards, the resin was been packed on the column. The unbound fraction was recovered and subject the resin to two wash steps: (1) 3 mL of lysis buffer, and (2) 3 mL of wash buffer (20mM imidazole, 50mM sodium phosphate dibasic and 300mM NaCl) which were collected separately. Finally, the proteins were eluted by using 3 mL of elution buffer (250mM imidazole, 50mM sodium phosphate dibasic and 300mM NaCl) collected in 6 different fractions of 500 µL.

5.2.2 *SDS - PAGE*

Once the purification was performed, 30 µL of each fraction underwent a solubilization process by adding SDS-PAGE Loading Sample Buffer 4X (1 M Tris-HCl pH 6.8, 8% SDS, 40% Glycerol, 5% β-mercaptoethanol, 10% Bromophenol blue), followed by a boiling step of 5 minutes. Hearafter, they were ready to be used for electrophoresis on polyacrylamide gel.

The denaturated proteins were loaded in a precast polyacrylamide gel (Bolt 4-12% Bis-Tris Plus 1.0 mm X 10 well, Invitrogen by Thermo Fisher Scientific). The electrophoretic run was carried out in MOPS buffer pH 7.5, containing: 50 mM MOPS (3-Morpholinopropane-1-sulfonic acid), 50 mM Tris-HCl, 1.025 mM EDTA (Ethylenediaminetetraacetic acid) and 0.10% SDS at a constant voltage of 100 V.

The proteins were then stained with Coomassie Brilliant Blue (1 g/L Coomassie Brilliant Blue, 50% methanol, 10% glacial acetic acid, 40% H_2O) and acquired by using the epi-illumination option of Amersham Imager 600.

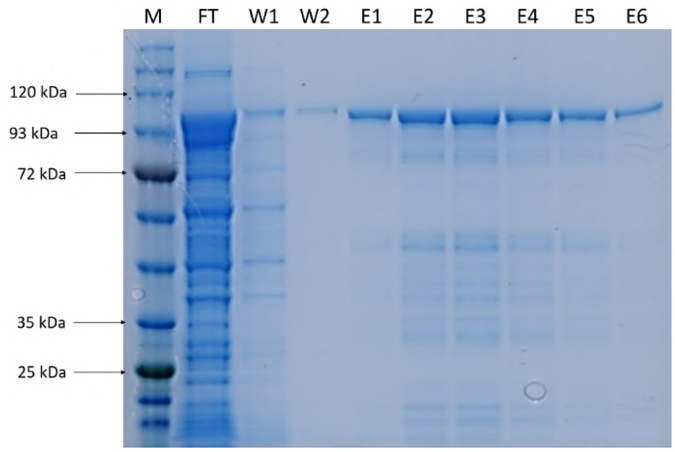

Figure 5.5.: **Purification of** $I27_4 - AS - I27_4$ **polyprotein:** M = protein marker Sharpmass VI, on the left the molecular weight of the reference bands (kDa) are reported, on the top of each lane is indicated the name of the fraction coming from the purification procedure: SP = total soluble proteins, FT = flow through, W 1-2 = wash, E 1-6 = eluted fraction. The solubilisation volume loaded in each lane is as follows: FT = 25 µL, W 1-2 = 20 µL, E 1-6 = 10 µL. The electrophoretic run has been conducted in a 4-12 % polyacrylamide gel, then subjected to Coomassie Blue stain.

The final polyprotein has a weight of ~100 kDa. In all of the samples there is a band at the same height as the one expected for the polyprotein, meaning that all were successfully expressed (See Fig.5.5 for a standard $I27_4 - AS - I27_4$ polyprotein purification).

In the lanes that corresponds to the flow through and the wash (FT and W) there is a band just above 100 kDa, which can be attributed to the polyprotein. This means that the quantity of the loaded proteins on the resin is higher than one that the resin can bind. Indeed, in all the last lanes (eluted fractions E1-E6) there is an intense band that corresponds to the polyprotein, along with some less intense bands that we observed in the Western Blot (Fig.5.6). The fractions E2 and E3, the most intense ones, were used for the SM-FS experiments.

5.2.3 *Western Blot and immunodecoration*

In the case of Western Blot essay, after the SDS-PAGE, the proteins were transferred from the gel to a 0.2 µm pore-size nitrocellulose membrane using a transfer buffer composed of 25 mM Tris Base, 190 mM glycine, 20% V/V methanol, 0.1% p/V SDS.

The wet transfer process was conduct at 4°C for 2 hours at a constant electrical field of 250 mA.

For the immunodecoration, in order to prevent non-specific background bindings, the nitrocellulose membrane was incubated in a

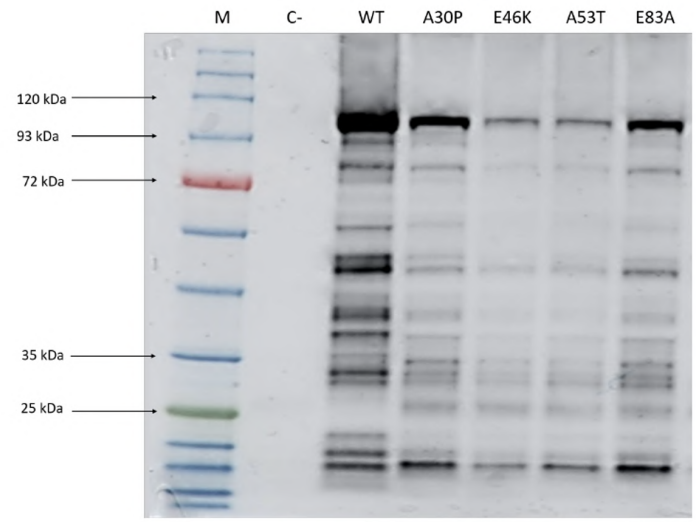

Figure 5.6.: **Determination of the expression levels of the polyproteins by Western Blot and Immunodecoration**: M = protein marker Sharpmass VI, on the left the molecular weight of the reference bands (kDa) are reported, on the top of each lane is indicated the name of the samples; C- = cells coming from non-transformed BL21 (DE3). For each sample 1 mL of cell culture was pellet, resuspended in 50 µL of sterile milliQ water and solubilised. For each sample, 15 µL of a 1:5 diluition of the solubilisation product, in SB1X, were loaded on a 4-12% polyacrylamide gel, transferred on a 0.2 µm nitrocellulose membrane and incubated with HRP-His probe 1:5000.

blocking solution, containing 2.5% BSA dissolved in T-TBS (0.05 % Tween-20, 25 mM Tris-HCl pH 7.5, 0.15 M NaCl), for one hour at room temperature under gentle agitation and then washed three times in T-TBS. Afterwards, the membrane was incubated with 0.8 mg/L antibody HRP-His Probe (Thermo Fisher Scientific, cat. 15165) in blocking solution for one hour and washed again three times in T-TBS. Finally, the presence of the proteins was tested by incubating the membrane with a 1:1 mixture of Luminol Enhancer solution and peroxide Buffer solution (EuroClone, cat. EMP012001) for 2 minutes in the dark. The final image was acquired by using the chemiluminescence option of Amersham Imager 600 for 5 to 30 seconds, depending on the signal intensity.

As identified also in the Comassie Blue stain essay (Fig.5.5), a ~ 100 kDa band was present in all the samples, as well as the truncated polyprotein signal at lower molecular weight.

However, the intensity of the $I27_4 - AS - I27_4$ polyprotein bands was different: the WT protein shows the most intense signal, meaning that it had higher expression comparing to the other mutants, the A30P and E83A mutants showed ~ 30% of the WT intensity, but were

more expressed than the other two mutants, that were less than ~ 10% with respect to the WT.

5.3 NATIVE MASS SPECTROMETRY

Nano-ESI-MS data were taken from [149].

5.3.1 *Sample preparation*

Briefly, Ammonium acetate, ammonium hydroxide, dopamine (DA), epigallocatechin-3-gallate (EGCG) were purchased from Sigma Aldrich (St. Louis, MO). Wild-type human AS was expressed in recombinant form and purified as described in [218] in the G. Legname lab. After sample desalting against 100 mM ammonium acetate pH 7.4 on PD-10 columns (Amersham Biosciences, Amersham UK) to promote sodium displacement, aliquots were lyophilized and stored at -80°Cuntil further use.

5.3.2 *Native MS*

The stock solution of EGCG (1 mM) was prepared in 10 mM ammonium acetate pH 7.4. The stock solution of DA (100 mM) was prepared in 10 mM ammonium acetate pH 5 and then diluted in 10 mM ammonium acetate pH 7.4 immediately before the experiments to reach the desired concentration. The final pH of the solutions was checked by the Crison Basic20 pH meter (Crison Instruments, Barcelona, Spain).nano-ESI-MS spectra were collected after 10 minute incubation of protein-ligand mixtures in 10 mM ammonium acetate, pH 7.4, at a final AS concentration of 20 μM. Nano-ESI-MS spectra were collected in positive-ion mode, using a hybrid quadrupole-time-of-flight mass spectrometer (QSTAR-Elite, Biosystems, Foster City, CA) equipped with a nano-ESI sample source. Deconvoluted spectra were obtained using the Bayesian protein reconstruct tool of BioAnalyst™extension of Analyst QS 2.0 software (Applied Biosystems, Foster City, CA). Quantification from native-MS data was based on Gaussian fitting of CSDs, upon transformation to x=z abscissa axis. The values reported refer to the area of the components obtained for the protein in the absence of ligand, for the 1:1 AS:ligand complexes and for all the stechiometry of AS:ligand complexes.

5.4 CD AND FTIR ANALYSIS

Circular Dichroism (CD) and Fourier-Transform Infrared Spectroscopy (FTIR) analyses were performed as described in [162] . In particular, Far-UV CD spectra of 20 μM AS in PBS buffer were acquired on a J-815 spectropolarimeter (JASCO Corp., Tokyo, Japan) under the following instrumental settings: data pitch, 0.1 nm; scan speed, 20 nm/min; bandwidth, 1 nm; accumulation spectra, 2. A 1-mm path length quartz cuvette was employed. FTIR spectra of 340

μM AS in deuterated PBS buffer were acquired on a Varian 670-IR spectrometer (Varian, Mulgrave VIC, Australia Pty. Ltd.) under the following instrumental settings: resolution, 2 cm^{-1}; scan speed, 25 kHz; scan coadditions, 1000; apodization, triangular; nitrogen-cooled mercury cadmium telluride detector. A temperature-controlled transmission cell with two BaF_2 windows separated by a 100-μm Teflon spacer was employed. At least data from three independent experiments were performed.

5.5 SMFS MEASUREMENTS CONDITIONS

5.5.1 *Fluid cell for SMFS*

A fluid cell was created by firmly attaching a glass o-ring on the gold-coated cover slip by using vacuum grease, obtaining a final volume of ~ 2 mL. The gold substrate has been extensively employed into SMFS of proteins because it can achieve the necessary strong attachment with one end of the poliproteins, by taking advantage of the covalent Thiol group formed between the gold and at least one cystein (our constructs carry 2 cystein). Glass cover slips (VWR, 24 mm ⌀, Microscope cover Glasses) were individually washed with detergent and they were sonicated 20 minutes in 100% pure 2-propanol (Sigma Aldrich, St. Louis, MO) and 20 minutes in milli-Q water. This cycle was repeated twice. Each coverslip was individually dried with nitrogen and a gold layer of aboute 20-40 nm was directly evaporated on the clean glass by means of a Gold Sputter Coater (Quorum technologies, SC7620 Mini Sputter Coater). Gold-evaporated coverslips were then placed individually into storage boxes and conserved in the dark into a vacuum-sealed container for at maximum one month. On the other side of the chain, the protein was attached to the silicon nitride cantilever through unknown interactions, that probably involved physical adsorption.

5.5.2 *Sample preparation for SMFS measurements*

SMFS experiments were carried out on a Nanowizard II (JPK Instruments, Berlin) at room temperature. Prior to each experiment, every cantilever (Si_3N_4, Bruker MLCT-BIO, Cantilever D, Nominal spring constant k=0.03 N/m, for more details about cantilevers calibration see Sec.1.3.4) was individually calibrated using the Equipartition Theorem in the JPK software. Approximately 20μL of protein (at a concentration of ~ 2μM) were deposited onto the evaporated gold coverslip and allowed to adsorb at room temperature for about 15 min without let the solution dry out completely. After this time, 1.8 mL of sterile PBS buffer (NaCl 8 g/L, KCl 0.2 g/L, Na_2HPO_4 1.42 g/L, KH_2PO_4 0.24 g/L, final pH 7.4, 150 mM) were added to reach an

overall protein final concentration of ~ 20 nM. Each experiment was carried out in fresh PBS buffer, to which EGCG (stock diluted in PBS) and DA (stock diluted in acidic MilliQ, pH 4) (stored at 4°Cprotected from light) were added to reach the desired final concentration (200 μM of DA, and 25 μM of EGCG). Each solution was filtered on a filter screen with a porosity of 0.2-μm before each experiment.

5.5.3 *AFM-based SMFS*

In a typical constant-velocity single-molecule pulling experiments, the protein, tethered between the tip and the sample surface, was stretched moving away the tip with a constant speed of 1μm/s with a recording rate of 4096 Hz. The laser beam displacement was converted to force and plotted against the tip-sample separation to obtain the final force curve.

Briefly, the tip approached the sample surface with a constant speed, pushing the surface with a fixed force for a fixed time (respectively, *Pushing Force* and *Pushing Time* in Tab.5.3) in order to promote the absorption to the tip of a single protein deposited on the surface. Then, the tip was pulled away from the surface with a constant speed and the deflection of the cantilever (i.e. the interaction force between the sample and the AFM tip) was recorded as a function of tip-sample separation.

In every experiment, thousands of curves were recorded along a square grid with a side of 10μm and 32 x 32 points. The approaching speed of the probe towards and backwards the sample was 1 μm/s, while the distance travelled by the tip was 0.8 μm, enough to have a complete detachment of the protein from cantilever.

Pushing Force	2 - 3 nN
Pushing Time	0 - 2 s
Pulling speed	1 μm/s
z - Range	800 nm
Rate of recorded points	4096 Hz

Table 5.3.: Range of values for SMFS pulling experiments.

5.5.4 *Data Analysis*

The force curves were firstly processed by both the JPK-Data Processed software (JPK Instruments, Berlin) and MATLAB custom-written software which was specially implemented. Briefly, the curves were corrected for the baseline and the deflection of cantilever by means of JPK-DM software. A first screening process was automatically exploited by searching curves presenting more than 6 peaks. Generally, during this operation, the 50-70% of curves were usually

discarded. The second selection of the curves, which was slightly user-dependent, discarded all the traces with spurious signal, misfolded peaks or more than one polyprotein attached. The selected curves were individually analysed, checking the goodness of both the baseline and the recorded peaks. The contour length (L_c) of each peak (both I27 and AS) was calculated by fitting with a single parameter WLC model (while the persistence lenght L_p = 0.36 nm was kept constant). Only the curves with a single clear detachment peak, 6-8 (7-9 for SI-like AS conformation, see Chapter 6 for details) peaks and no noise in the first part were considered.

The final criteria for the acceptance of the curves for further analysis were:

- The contour length relative to the first peak refereed to the AS had to be in the range of 80 $\pm$ 15 nm in the case of *Random Coil* (RC) or *Mechanically Weak Interactions* (WI) conformations, and 40 $\pm$ 15 nm for the *Mechanically Strong Interactions* (SI).

- The total number of peaks for the polyprotein containing AS in RC or WI conformations had to be from 6 to 8 and from 7 to 9 in the case of SI.

- The detachment peak had to be clean, well defined and occurring at a force higher than the one of the I27 denaturation.

- Non-specific interaction (spurious) signal at the beginning had to not exceed the first 25 nm and their height had to be lower than the other peaks.

- The distance between two consecutive peaks had to be compatible with the length of the unfolded I27 modules and their height with the values reported in the literature ($\sim$ 260 pN for I27 at a pulling speed of 1 μm/s).

These criteria, very strict and conservative, led us to discard a discrete number of curves that shown a clear AS signal but with some problems in a variable part of the trace.

By employing the JPK-DP software the range of each single peak was manually controlled and fitted by means of WLC model. The last peak, which describes the detachment, was always discarded for the analysis. The force of unfolding F_U was found as the maximum force of the peak (i.e the highest point of the peak), while the contour length (L_c) was obtained as free parameter from the WLC fit. Also the length size of the linkers between two proteins should be considerate to not overestimate the protein length. Despite this, it has been demonstrated that the linkers are totally extended before the unfolding [73].

Further statistical analysis had been done by using different self-written codes in MATLAB language.

6

RESULTS: NANOMECHANICAL UNFOLDING OF AS IN THE PRESENCE OF LIGANDS OR MUTATIONS

The SMFS experiments have been performed on a polyprotein construct containing eight repeats of titin immunoglobulin-like domain I27 (Section 2.2.2) and one grafted AS domain (Fig.6.1). The I27 domains act as molecular handles to mechanically stretch the AS protein and introduce a well-characterized fingerprint signals into the recorded force signal that make possible the identification of the different AS conformations.

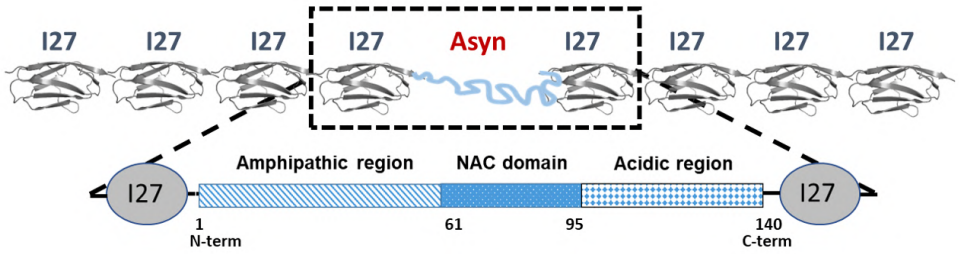

Figure 6.1.: Polyprotein construct encompassing the AS full-length polypeptide chain for SMFS experiments.

6.1 CONFORMATIONAL ENSEMBLE OF AS

Three distinct classes of conformational structures in equilibrium were identified by using the nanomechanical unfolding in near physiological condition (PBS buffer, pH 7.5, 150 mM): Random Coil (RC), collapsed states stabilized by Weak Interactions (WI), and compact conformations stabilized by Strong Interactions (SI). A representative sketch of the AS in the three conformations is reported in Fig.6.2.

As shown in Fig.6.3, the observed SMFS curves show the typical *sawtooth* pattern in which each peak can be assigned to the unfolding of an individual I27 protein. The initial part of the curve, instead, correlates with the presence of AS and it is characterized by different mechanical resistances to unfolding. These AS unfolding pathways, which are described in details below, are consistent with the presence

Figure 6.2.: Examples of AS structural conformations as obtained in SMFS. RC (blue), WI (green) and SI (red). The last 45 residues (N-term portion) are always in a not-structured form.

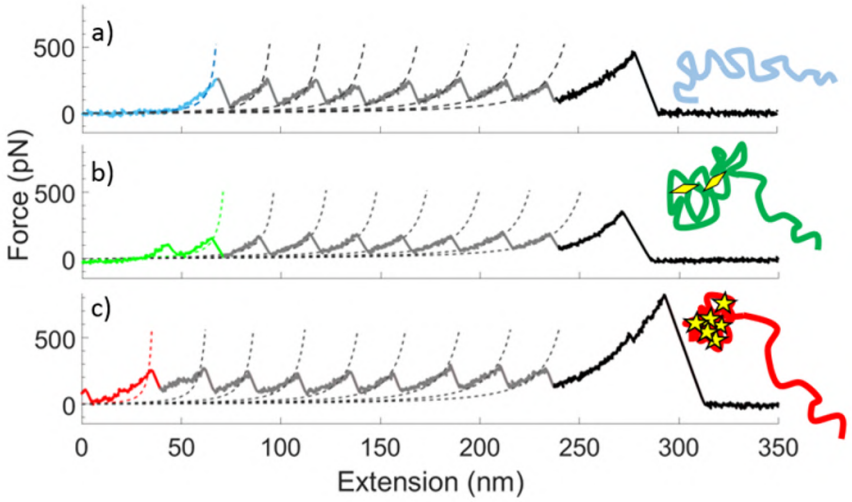

Figure 6.3.: Representative force curves of the mechanical unfolding of the polyprotein in distinct conformations stabilized by RC (a), WI (b), and SI (c). Figure adapted from [1].

of an AS heterogeneous conformational ensemble, already discuss in literature [88, 89, 169]. Every curve was fitted by means of the WLC model to extract the contour length L_C of each peak (both for I27 and AS). The three distinct patterns can be recognized by analyzing the L_C of the first peak (Fig.6.3 and Fig.6.4).

The first class of curves showed the first peak after L_C=79±6 nm (light blue curve, first line of Fig.6.3), followed by the typical sawtooth pattern relative to other I27 consecutive unfolding events. This curves were ascribed to the unstructured conformation of AS and classified as RC, since no additional peak is detected in the first ~80 nm, showing a complete elongation of both the folded polyprotein and of the 50 nm-long AS. In this kind of event, the AS was captured in a conformation whose complete extension occurred at low force and did not imply the overcoming of measurable unfolding energy barriers. This behaviour is typical of random coils and mostly unfolded conformers, and, unfortunately, only by SMFS measurements it is impossible to distinguish between them. Therefore, we can assume that this type

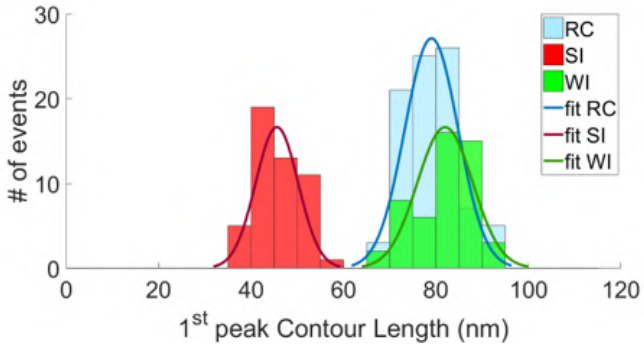

Figure 6.4.: Statistical distribution of the contour length of the first peak for RC (L_C=79±6 nm), WI (L_C=82±6 nm) and SI (L_C=46±5 nm) conformations for $I27_4$-AS-$I27_4$ polyprotein unfolding. Solid lines represent the Gaussian fits of the histograms. Figure adapted from [1].

of curves correspond to the AS in random coil state, identical to the one found by other techniques.

The second conformation was characterized by L_C=82±6 nm and by the presence of one-to-three small peaks before the first regular peak attributed to an I27 unfolding (green curve, second line of Fig.6.3). This pattern was associated to a collapsed conformation of AS stabilized by weak interactions (significantly smaller than the secondary-structure dentaturation signals) a and classified as WI, characterized by a smaller energy barrier to overcome.

The WI state signal, whose small (one to three) peaks correspond to an unfolding force (F_{WI}=117±34 pN) sensibly lower than I27 (F_{I27}=257±46 pN), see Fig.6.5, was characterized by a smaller energy barrier to overcome to unfold the AS.

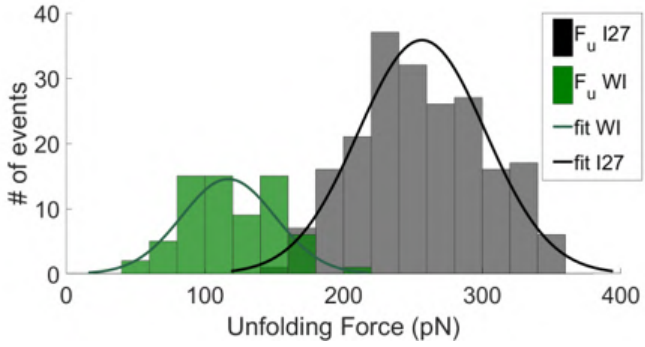

Figure 6.5.: Unfolding force statistical distribution of WI (F_{WI}=117±34 pN) and I27 modules (F_{I27}=257±46 pN). Figure adapted from [1].

The third conformation displays the first peak unfolding at the L_C=46±5 nm (red curve, third line in Fig.6.3). These curves were characterized by a shorter L_C and a comparable unfolding force for both AS and I27, and were assigned to a collapsed state of AS, mainly

stabilized by SI, which present a resistance to unfolding similar to the one of the high mechanostable β-rich protein I27.

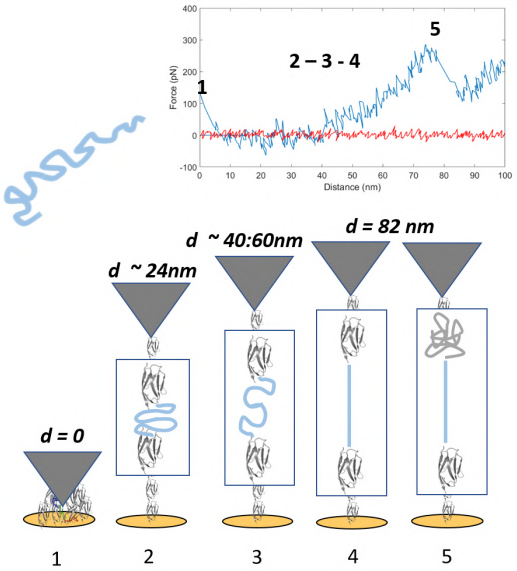

Figure 6.6.: **Random coil**. Schematic representation of the mechanical unfolding events that lead to the formation of the first peak for RC conformations.

In details, the length at which the first peak unfolds (in the curves assigned to the RC conformation, Fig.6.6) and that of the first of the higher peaks in the curves assigned to the WI (Fig.6.7) were all around 80 nm. These peaks occurred when AS was completely extended and flanked by eight I27 folded modules. The measured length was due to the contribution of the eight folded I27 modules (a folded module of I27 is 3 nm long, i.e. 3 nm x 8 = 24 nm), the length of eight linkers between each protein module (a linker is 2 aa, i.e. 8 x 0.36 nm x 2 = 5.76 nm) as already shown [219], and the length of the completely extended AS (i.e. 140 aa x 0.36 nm = 50.4 nm). By summing all the contributions, the total extension was 80.16 nm which was coherent with the measured values. The subset of curves presenting a L_C of the first peak higher than 95 nm, which could be associated with an undesirable misfolding event of a I27 module, was discarded.

The length of the first peak of the SI conformation (Fig.6.8) was due to the contribution of the eight folded I27 modules (3 nm x 8 = 24 nm), the length of eight linkers between each protein module (8 x 0.36 nm x 2 = 5.76 nm) as already shown [219], and the length of the completely extended N-domain of the AS protein (i.e. 45 aa x 0.36 nm = 16.2 nm). By this prediction, the first peak should occurs at 45.92 nm, which is compatible with the found value of L_C=46±5 nm.

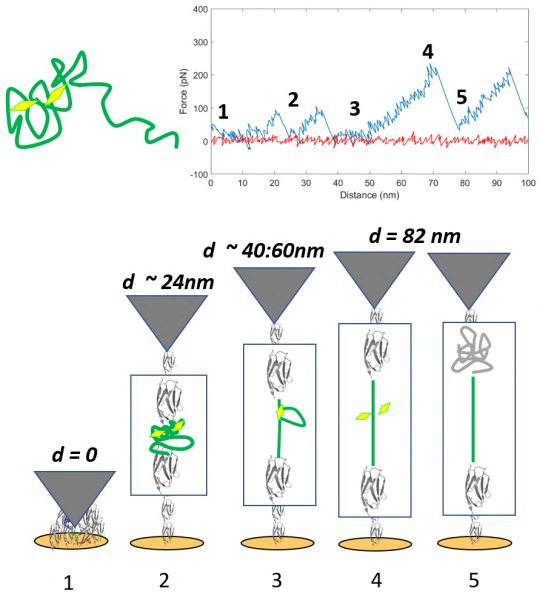

Figure 6.7.: **Weak Interactions** Schematic representation of the mechanical unfolding events that lead to the formation of the first two peaks for WI. In the detailed image of the first unfolding peaks, the two smaller peaks (# 2 - # 3) could be noticed. The position and the force relative to these peaks were F_{WI}=117±34 pN and $L_C(WI)$=41±14 nm).

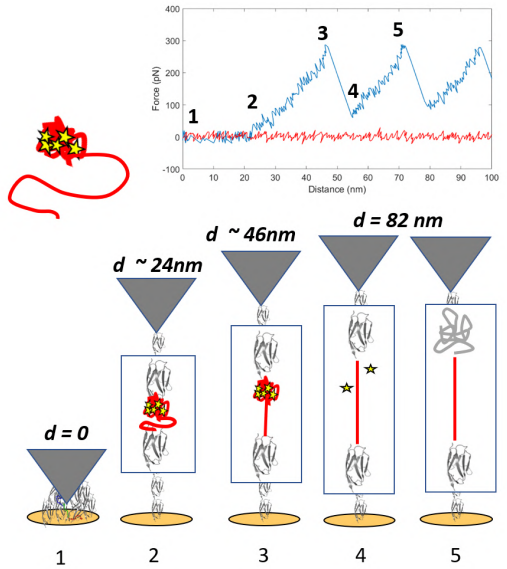

Figure 6.8.: **Strong Interactions** Schematic representation of the mechanical unfolding events that lead to the formation of the first two peaks for SI. The first peak has the same height as the other peaks and occurs after 46±5 nm. The average L_C of the second peak is 80±9 nm.

After having characterized the three different conformations of AS after the nanomechanical unfolding, several session of SMFS had been performed without ligand, with the presence of EGCG and DA, at different pH and with three of the known point mutation associated with the insurgent of PD (A30P, A53T and E83A).

6.2.1 *Standard conditions*

The first session of measurement was performed in PBS (pH 7.4 and ionic strenght of 150mM). This first step was necessary to study the conformational states of AS in a condition that resembles the physiological condition, with an osmolarity and ion concentrations analogous to the ones that could be found in the human cells. After a very restrictive curve selection, a total of 60 curves were considered, with a prominent preference for the unstructured form of AS (RC) and repartitioned as shown in Tab. 6.1.

	AS	
	# of curves	% of curves
Random coil	37	62±6
Strong Interactions	5	8±5
Weak Interactions	18	30±4
N(tot)	60	100%

Table 6.1.: Number of curves considered for the construct $I27_4$-AS-$I27_4$ for WT AS in absence of ligands. Error in percentage column represent the standard deviation calculated for the normal distribution.

These data might be compared to the one reported in [88] (RC 38.2%, SI 7.3% and WI 54.5% for AS in Tris 10mM or RC 49.2%, SI 27.9% and WI 22.9% for AS in Tris 500mM). However, it had been reported that for the IDPs, and consequently for AS, slightly differences in the environmental conditions may drastically influence the conformation of this protein. *Hervás et al.*, in [169], divided the AS behaviour into two classes: Non Mechanostable (NM), characterized by a forces lower than 20 pN and Mechanostable (M) for larger unfolding forces. For AS in standard conditions the conformations were allocated in 55.1% NM and 44.9% M in PBS pH 7.4. This results was compatible with our finding, if considering this conversion: RC $\rightarrow$ NM and WI+SI $\rightarrow$ M.

6.2.2 *DA*

It has been reported that AS changes its structure in response to DA, adopting a conformation where its N- and C-terminus become closer together, thus, inhibiting AS fibrils formation and leading to the accumulation of AS oligomeric species via an alternative folding pathway [220]. Native-MS based studies had shown that DA binding to AS caused a tendency to rearrange the complex in a structured thus still partially compact state [149]. In order to compare the AS structural conformation in the presence of DA, avoiding technical problems, a DA concentration of 200μM was chosen. After the selection, a total of 73 curves were considered, with a prominent preference for the unstructured form of AS (RC) and repartitioned as shown in Tab.6.2.

AS with DA 200 μM

	# of curves	% of curves
Random coil	31	42±5
Strong Interactions	23	32±5
Weak Interactions	19	26±5
N(tot)	73	100%

Table 6.2.: Number of curves considered for the construct $I27_4$-AS-$I27_4$ for WT AS in the presence of 200μM DA. Error in percentage column represent the standard deviation calculated for the normal distribution.

Comparing these results with the AS in absence of ligand, it was immediate to notice a decrease in the unstructured conformation and a strong increase in the SI, while the variation of WI conformation was not significant, falling inside the error uncertain (Fig.6.9).

6.2.3 *EGCG*

EGCG had been demonstrated to have a neuroprotective role in Parkinson's Disease (PD) on the basis of its anti-oxidative, anti-inflammatory, and fibril-destabilizing effects. Indeed, EGCG was capable of modulation the misfolding of proteins by directly binding to unfolded polypeptide chains and forming unstructured, nontoxic oligomers instead of the ordered β-sheet rich amiloyds fibrils. Moreover, it is known that EGCG is able to reduce the structural stability of AS toxic fibrils and remodel them in a more disordered and less toxic oligomeric structure [213]. Native-MS based studies had shown that EGCG binding to AS caused a tendency to rearrange the complex in a compact structure [149]. A total of 53 curves were collected in presence of EGCG (Tab 6.3) a remarkable change in the percentage of the three population with respect the same experiment in the absence of ligands could be noticed.

AS with EGCG 25μM		
	# of curves	% of curves
Random coil	19	36±6
Strong Interactions	21	40±7
Weak Interactions	13	24±6
N(tot)	53	100%

Table 6.3.: Number of curves considered for the construct $I27_4$-AS-$I27_4$ for WT AS in the presence of 25μM EGCG. Error in percentage column represent the standard deviation calculated for the normal distribution.

Indeed, there was an increase in the number of AS secondary-like structure SI, that increased from 8% to 40%. Accordingly, the percentage of AS in the unstructured state decreases from 62% to 36% in the presence of EGCG (Fig.6.9).

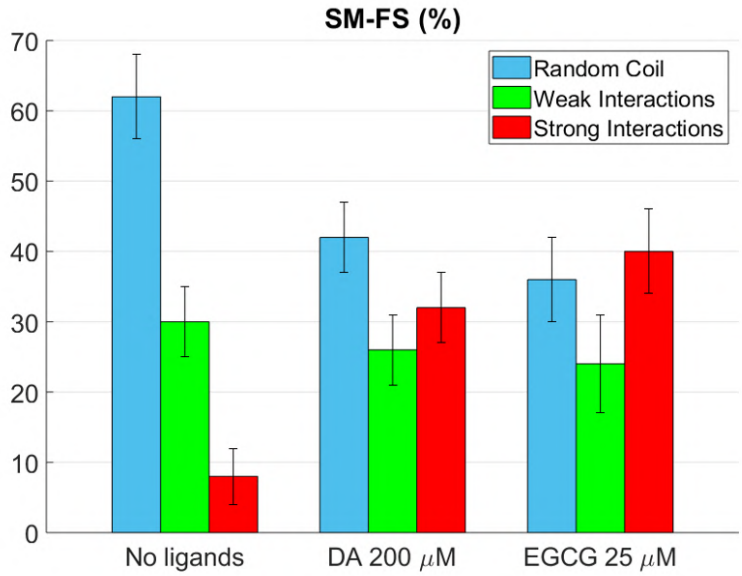

Figure 6.9.: **Species distributions in the presence of Ligands** Different population of AS conformers obtained by SMFS in absence of ligands (Right), in the presence of 200μM of DA (Middle) or 25μM EGCG (Left). Error bars represent the standard deviation calculated for the normal distribution. Figure adapted from [1].

6.2.4 *Mutants*

The mutants related to the insurgence of PD showed to drive the AS protein along different fibrillation and oligomerisation pathways. For instance, the single point mutations A30P and A53T, located in the amino-terminal sequence of AS, seemed to have effects on variation

in protein membrane interaction or in protein aggregation propensity [107, 109]. The quite recent mutation, E83A, discovered in the core region of AS NAC domain, had been reported to favour the β-structures involved in fibril formation and aggregation [113]. Therefore, with the aim of studying the effects of these mutations on the conformational equilibrium of AS, we produced polyproteins containing the AS point mutations A30P, A53T and E83A. Preliminary results are reported in Tab. 6.4 and Fig.6.10.

AS Mutations						
	A30P		A53T		E83A	
	#	%	#	%	#	%
Random coil	14	41±8	11	44±8	12	44±8
Strong Interactions	5	15±6	4	16±6	7	27±5
Weak Interactions	15	44±8	10	40±8	7	27±5
N(tot)	34	100%	25	100%	26	100%

Table 6.4.: Number of curves considered for the construct $I27_4$-AS-$I27_4$ for A30P AS, A53T AS and E83A AS in PBS pH 7.4. Error in percentage column represent the standard deviation calculated for the normal distribution.

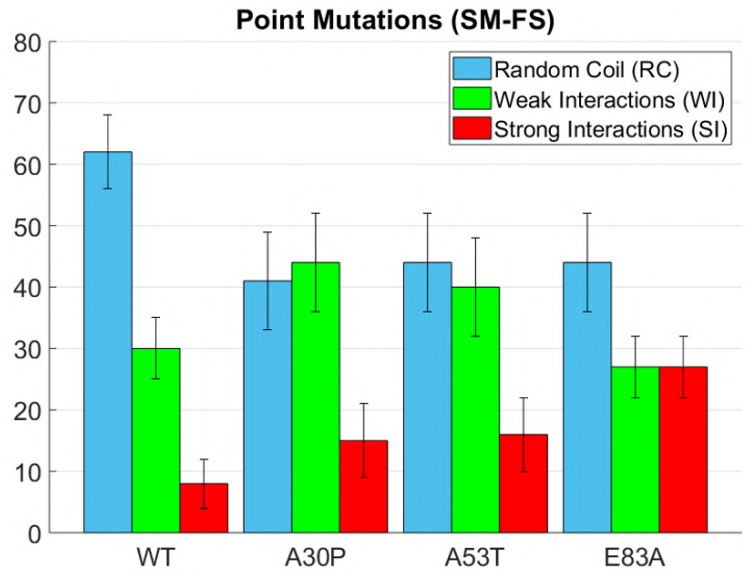

Figure 6.10.: **Species distributions for point mutation AS** Different population of AS conformers obtained by SMFS for (from right to Left) WT-AS, A30P-AS, A53T-AS and E83A-AS. Error bars represent the standard deviation calculated for the normal distribution.

6.3.1 *Characterization of the I27 module: contour length and unfolding force*

As a consequence of the unfolding of our chimeric polyprotein $I27_4$-AS-$I27_4$, the several protein I27 module were unfolded under the mechanical action of the applied pulling force. As a result, a series of peaks in the force-extension curve have been detected. Each peak was individually analysed considering its maximum force value (F_U), which corresponds to the force necessary to unfold a single I27 module. Furthermore, the force-extension data were fitted with the WLC model in order to extract the asymptotic value (L_C) of the extension of the each I27 module. Being the I27 module well characterised in literature [17, 74, 79, 207], the study of the statistical distribution of the modules both in L_C (28 nm) and F_U ($\sim$ 270 pN at v=1μm/s) was a good indicator of a proper calibration.

In Tab.6.5 and Fig.6.11 the statistical distributions of the F_U and L_C values for I27 as obtained in the different conditions (in PBS, in 200 μM DA, and in 25 μM ECGC) were reported. As shown by the reported histograms (Fig.6.11), there was no appreciable difference between the F_U and L_C behaviour of I27 changing from PBS to DA or EGCG. These results clearly indicates that the DA and EGCG alone were not capable to alter the mechanical stability of a non-specific and structured protein, such as the I27.

I27 characterisation F_U and L_C

	PBS	200μM DA	25μM EGCG	Total
F_U - I27 (pN)	274 ± 42	227 ± 40	271 ± 39	257 ± 46
L_C - I27 (pN)	28.2 ± 1.1	28.0 ± 0.7	28.1 ± 0.9	28.1 ± 0.9

Table 6.5.: Measured mean values $\pm$ standard deviation of the unfolding force F_U and of the average contour length L_C measured for the I27 module for each force-extension curve. The values reported in the last column represent the cumulative statistical values of data acquired both in the presence or absence of ligands.

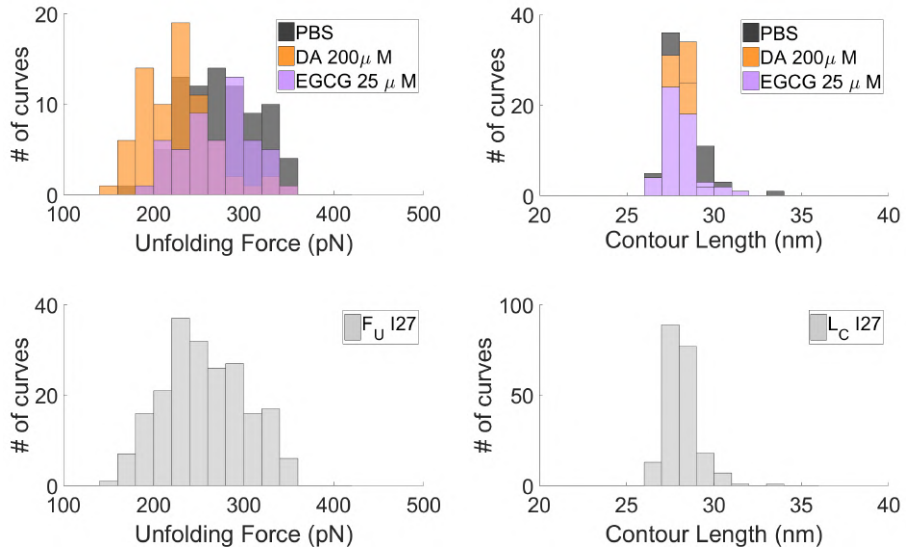

Figure 6.11.: Statistical distribution of the average unfolding force F_U (histograms in the left column) and of the average contour length Lc (histograms in the right column) measured for all the I27 module, for each force-extension curve. Histograms in the first line represent data obtained in PBS, 200 μM Dopamine (DA), and 25 μM Epigallocatechin gallate (ECGC). Histograms in the second line represent the cumulative statistical distribution of data shown in the first line. Figure taken from [1].

6.4 FIRST PEAKS IN THE FORCE-EXTENSION CURVES: CONTOUR LENGTH AND UNFOLDING FORCE

In the Fig.6.4 the histograms of the contour length L_C of the first peak of each curve were reported as cumulative for the three tested WT-AS condition (alone or with DA or EGCG ligands). In Tab.6.6 the contour length values (mean ± std) of the first peak for the three conformations in PBS, 200 μM DA and 25 μM EGCG separately were reported. The last column (*Total*) reports the cumulative L_C values.

1st peak contour length L_C (nm)

	PBS	200μM DA	25μM EGCG	Total
Random Coil	79.1 ± 5.7	78.7 ± 5.2	79.5 ± 7.0	79.0 ± 5.8
Strong Interactions	40.7 ± 1.2	47.0 ± 3.9	45.1 ± 5.0	45.6 ± 4.6
Weak Interactions	82.0 ± 5.8	81.9 ± 5.8	82.1 ± 6.9	82.0 ± 6.0

Table 6.6.: Measured average values ± standard deviation of the contour length of the first peak for each force-extension curve in the different conformation state acquired in PBS, DA, EGCG. The values reported in the last column represent the cumulative statistical values of the data acquired in different buffer conditions (PBS, DA, EGCG).

Similarly, in Tab.6.7 we reported the unfolding force values (mean $\pm$ std) measured in correspondence of the first peak for the three conformations (RC, SI and WI), in PBS, 200 μM DA and 25 μM EGCG, separately. The last column *Total* reports the cumulative F_U statistical values, regardless of the different buffer conditions (PBS, DA, EGCG). The values reported in Tab.6.6 and 6.7 confirmed that ligands did not affect the contour length and the unfolding force of the first peak, even if they induced changes in the statistical distributions of conformations.

1^{st} **peak Unfolding Force** F_U (pN)				
	PBS	200μM DA	25μM EGCG	Total
Random Coil	221 ± 34	187 ± 37	215 ± 45	207 ± 40
Strong Interactions	208 ± 26	195 ± 50	217 ± 32	206 ± 42
Weak Interactions	218 ± 40	196 ± 36	234 ± 33	214 ± 39

Table 6.7.: Measured average values $\pm$ standard deviation of the unfolding force of the first peak for each force-extension curve in the different conformation state acquired in PBS, DA, EGCG. The values reported in the last column represent the cumulative statistical values of the data acquired in different buffer conditions (PBS, DA, EGCG)

In Fig.6.12 the histograms of both the first peak unfolding force (column on the left), and contour length (column on the right) for the three conformations (RC, SI and WI), in PBS, DA and EGCG were analysed separately. From these data it is possible to conclude that the unfolding force value and the contour length of the first peak were not affected by the presence of the ligands.

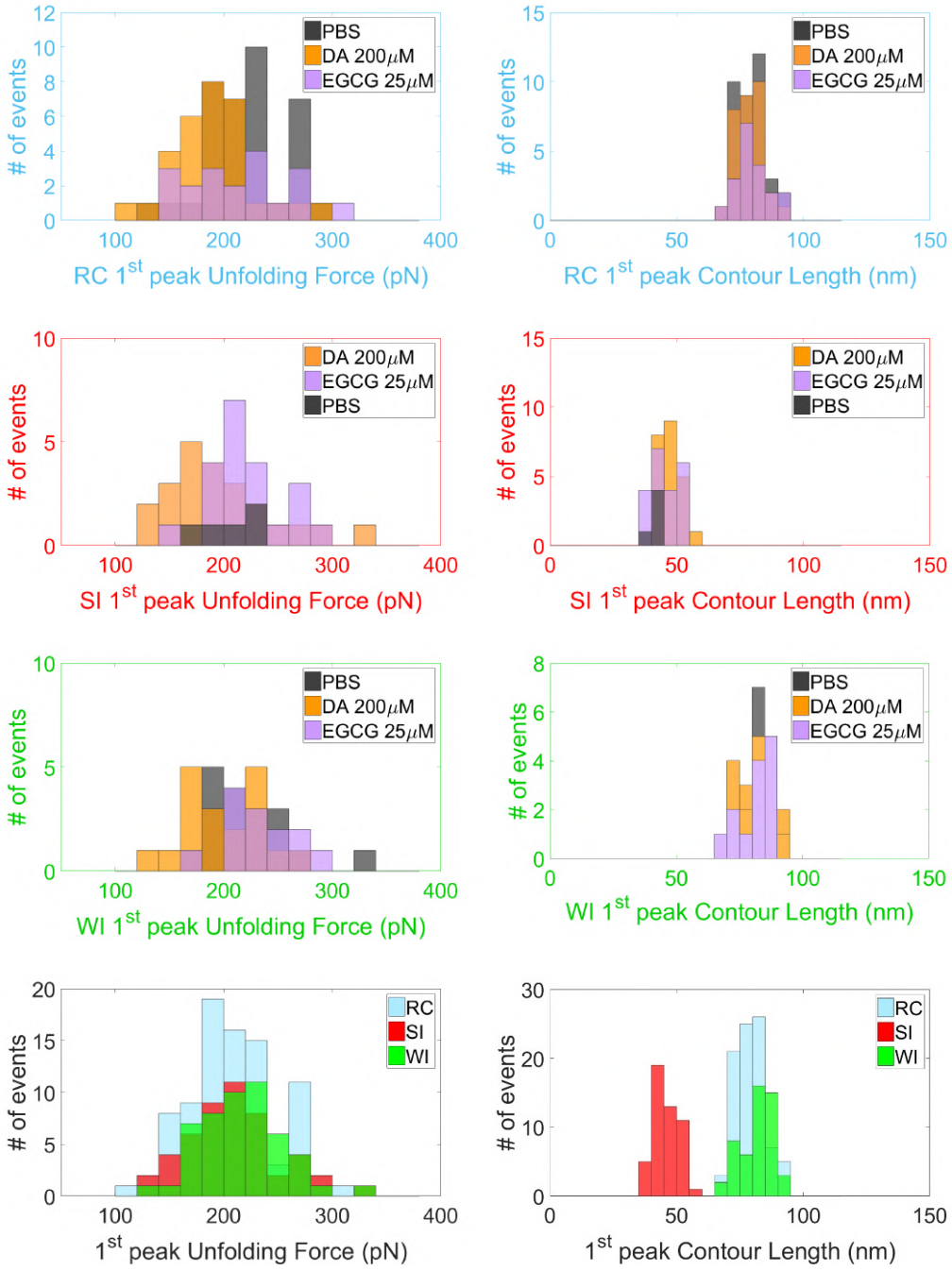

Figure 6.12.: Statistical distribution of the unfolding force F_U (in the left column) and of the contour length L_C (in the right column) of the first peak in each force-extension curve. Data for RC, SI and WI divided depending of their environmental condition (line 1-3) and represented cumulatively (last line). Figure taken from [1].

In Fig.6.5 the unfolding force of WI peaks were compared with the distribution of unfolding force of I27 module. In Fig.6.13 the unfolding force of the WI peaks were compared to the corresponding first higher peak.

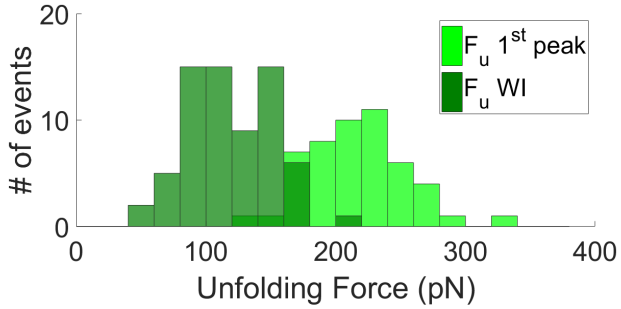

Figure 6.13.: Statistical distribution of the unfolding force F_U of the mechanically weak interactions (dark green) compared to the unfolding force F_U of the corresponding first peak collectively evaluated (PBD, DA and EGCG) in light green. Figure taken from [1].

In the Tab.6.8 and Fig.6.14, the statistical distribution of the F_U and L_C values of the WI as obtained in the different conditions (PBS, 200 µM DA, and 25 µM ECGC) were shown. As indicated in the figure, there were no appreciable difference between the WI behaviour of F_U and L_C in PBS, DA or EGCG.

1^{st} **peak Unfolding Force** F_U (pN)				
	PBS	200µM DA	25µM EGCG	Total
L_C(WI)-nm	43 ± 19	41 ± 11	39 ± 12	41 ± 14
F_U(WI)-pN	111 ± 34	108 ± 37	136 ± 19	117 ± 34

Table 6.8.: Measured average values ± standard deviation of the unfolding force and the averaged contour length measured for WI peaks in each force-extension curve. The values reported in the last column represent the cumulative statistical values of the data acquired in different buffer conditions (PBS, DA, EGCG)

Some WI traces showed more than one peak (typically 1 to 3), occurring in the range 30-65 nm. In order to understand if the number of peaks, their length and unfolding force were somehow correlate, we realized scatter plot of the unfolding force F_U of all the WI peaks versus their extension L_C (Fig.6.15). The histograms show separately the values obtained in different buffer conditions (PBS in black, 200 µM DA in orange, and 25 µM ECGC in purple). Data in Fig.6.15 took in account the presence of one (circles), two (squares) or three (diamonds) peaks in each force-extension curve, each peak contributed

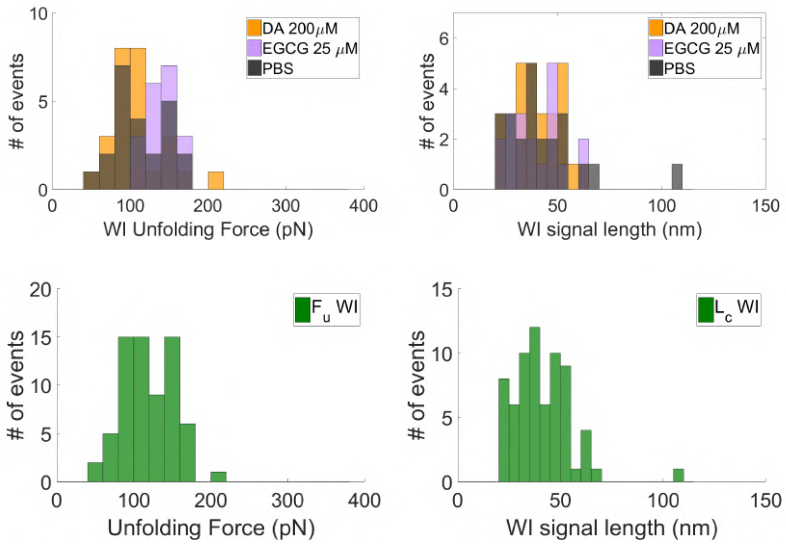

Figure 6.14.: Statistical distribution of the average unfolding force F_U (left column) and of the average contour length L_C (right column) measured for the WI peaks in each force-extension curve classified as WI. The histograms in the first line represent data obtained in PBS, 200 μM DA, and 25 μM ECGC. Histograms in the bottom line represent the cumulative statistical distribution of the data shown in the first line. Figure taken from [1].

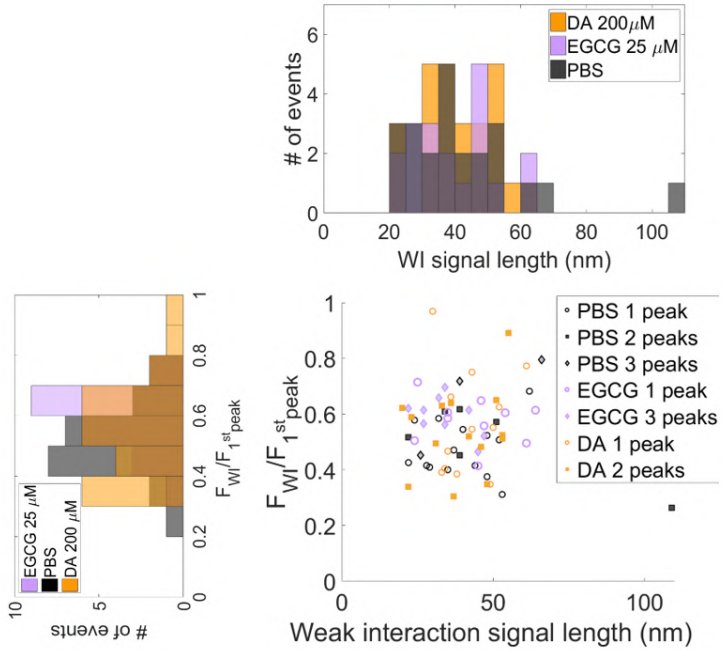

Figure 6.15.: Scatter plot of the unfolding force of the mechanically WI peaks and their extension length. Each dot represents a single WI unfolding peak (PBS in black, 200 μM DA in orange, and 25 μM ECGC in purple). Data took into account the number of unfolding peaks in each curve: a curve with a single peak contributes with one circle, a curve with two peaks contributes with two squares, a curve with three peaks contributes with three diamonds.

equally to the scatter plot. The graph shows no correlation between the two parameters.

6.6 CHARACTERIZATION OF THE SI PEAKS

The curves belonging to the SI class showed a first unfolding peak after ~46 nm, while all the following unfolding signals were not distinguishable. This was a limitation of the construct itself, as a matter of fact each peak was followed by an elongation of one single protein module. As already proposed in [88, 89] and considering the evidence that the last residues of the AS chain (from the 96^{th}-140^{th}) had never been observed into an ordinate and structured conformation, we proposed that the only initial AS portion (N-term + NAC domain, see Fig.6.1) was rearranged into a compact state stabilized by strong interactions.

Taking into account this hypothesis, the AS portion undergoing the structured form was formed by the first 96 amino acids, with an expected total elongation of 95x0.36 = 34.2 nm, but, considering a folded AS conformation of ~ 3 nm in length, the increment after the SI peak should be ~ 31 nm. This value was slightly larger than the one expected for the increment of an I27 module (89aa long, ~ $31.5 - 3 =$ 28.5 nm).

Some additional analysis have been performed to the SI curve, trying to determine if it could be possible to distinguish between the AS-SI peak from a I27 signal (Fig.6.16).

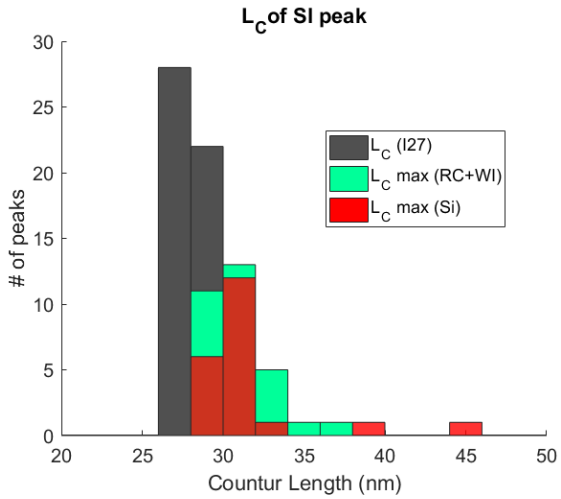

Figure 6.16.: Histograms of the mean L_C(I27) in grey, the maximum L_C increment for the SI traces (red) and of maximum L_C increment for the other two conformations (RC+WI, in emerald). Bin of the histogram 2 nm.

For each curve, the peak with the larger increment was compared to the mean of the other peaks in the same curve. This analysis was

performed also on the RC and WI traces for the same batch of curves, which in this case are used as a control.

Max L_C increment for SI			
	I27	max L_C SI	max L_C RC+WI
L_C (nm)	27.7 ± 1.2	31.0 ± 3.8	30.9 ± 1.9
N curves	53	21	32

Table 6.9.: Measured average values $\pm$ standard deviation of the mean L_C (second column) and max L_C increment for SI curves (third column) and RC+WI configurations (last column).

Unfortunately, as shown in Tab.6.9, even if the standard deviation of the L_C(SI) was sensibly higher than the one found in the control (RC+WI), the two mean values were still compatible within the errors. In conclusion, it was not possible to clearly assign a peak to the AS -SI unfolding. This was due principally to the variability to the calibration, as well as both the set up and analysis software resolutions.

6.7 OXIDATIVE LEVEL OF $I27_4$-AS-$I27_4$ POLYPROTEIN BY MASS SPECTROMETRY

To rule out that the conformational differences of $I27_4$-AS-$I27_4$ in the presence of DA or EGCG might be promoted by ligand-induced oxidation rather than ligand binding, the oxidative level of the polyprotein had been analyzed by MS-based proteomics techniques. the polyprotein was incubated 4 hours in the absence of ligands and in the presence of either 200 μM DA or 25 μM ECGC, in order to mimic the conditions of SMFS experiments. Then, a fast digestion of the polyprotein was performed by adding trypsin (protease:substrate 1:50) and incubating 30 minutes at 37°C. The reaction was stopped by the addition of 1% formic acid, and the tryptic peptides were desalted by C18 Ziptip (Millipore) before injection into an Orbitrap Fusion mass spectrometer coupled to a nano-HPLC system (EASY-nLC 1000, Thermofisher). Peptides were separated on a C18 column (length 500 mm, ID 75 μm, particle size 2 μm) by a 90-minutes gradient, analyzed by MS/MS experiments and identified by the software Proteome Discoverer (Thermofisher), using methionine oxidation as a variable modification. The sequence coverage was at least 92% for each sample run. The I27 domain contains one methionine residue (Met67). Wild-type AS contains four methionine residues (Met1, Met5, Met116 and Met127). However, Met1 has been removed in the construct employed in this work and the C-terminal ones, at positions 116 and 127, are known to be missed in a too large tryptic peptide under these conditions [162]. Thus, I27 Met67 and AS Met5 were used to monitor the oxidation level of the polyprotein. The extent of oxidation was quantified by the number of peptide-spectrum

matches (PSMs) relative to the total PSMs for the oxidized and non-oxidized variants of the corresponding peptides. Both methionine residues featurelow oxidation levels after the incubation (below 10% for AS Met5), and no induction by the ligands (Fig.6.17). This is in line with the evidence that DA- or EGCG-induced AS oxidation under similar conditions takes place on the time scale of days [162]. Furthermore, it has been shown that the oxidation kinetics of all AS Met residues is similar [162]. Thus, these results confirm that the effects of DA and EGCG revealed in this work by SMFS are ascribable to ligand binding DA or EGCG.

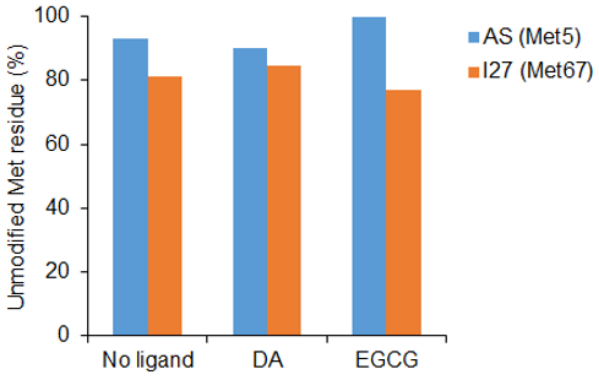

Figure 6.17.: Percentage of non-oxidized methionine residues after 4h-incubation of $I27_4$-AS-$I27_4$ polyprotein in the absence of ligands and in the presence of DA or EGCG. Figure taken from [1].

6.8 NATIVE MASS SPECTROMETRY

The statistical distributions of AS conformations alone or in the presence of ligands obtained by native MS showed the presence of intermediate compact states (I1 and I2) together with random coil (RC) and compact (C) conformations (Fig.6.18 for the 1:1 protein:ligand complexes and Fig.6.19 the cumulative MS data). The concentration of EGCG considered was not exactly the same for the two techniques (25 μM for SMFS and 80 μM for native MS) due to the limitation of both the instruments (lower concentrations were difficult to clearly detect in native MS and higher concentration in SMFS lead to interference with the AFM laser beam). The concentrations of DA was exactly the same (200 μM).

A quantitative comparison between the species distributions obtained by SMFS and native-MS data is shown in Fig.6.20. An intrinsic difference between SMFS and MS concerned the discrimination between free and ligand-bound protein molecules, which was possible only by the latter technique. Thus, native-MS data had been processed by two alternative ways. In one case, only signals of the 1:1

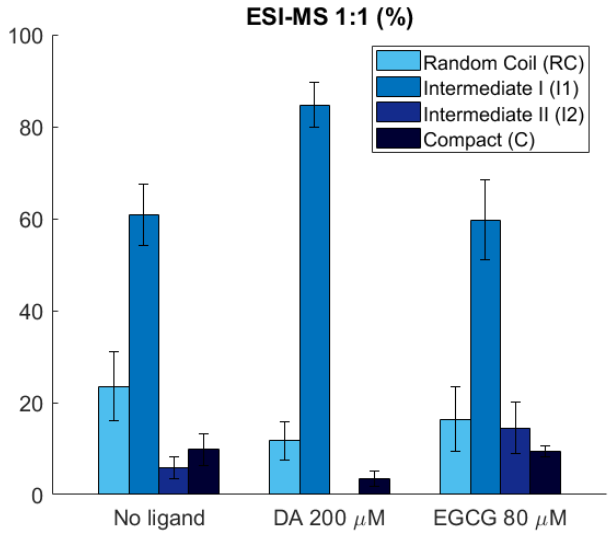

Figure 6.18.: Species distributions obtained by native MS (considering the 1:1 protein:ligand complexes); in the legend, between brackets, the intensity-weighted average charge of the peak envelopes is reported (i.e. RC=17.3; I1=13.7; I2=10.5; C=8.4). Error bars represent the standard deviations from three independent experiments. Data from [149]. Figure adapted from [1].

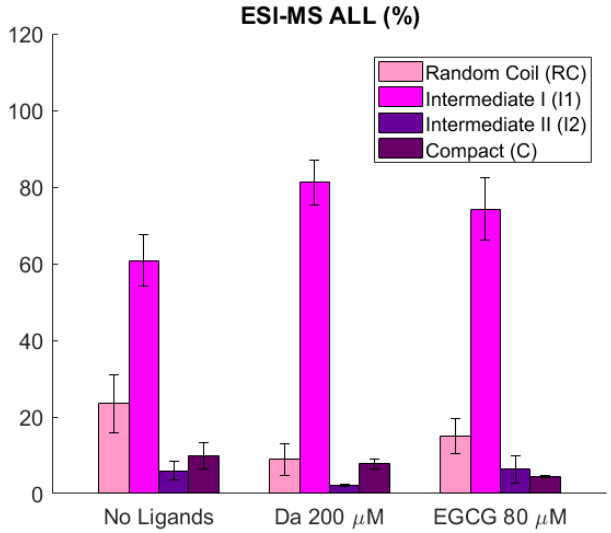

Figure 6.19.: Species distributions obtained by native MS (considering the cumulative MS data). In the legend, between brackets, the intensity-weighted average charge of the peak envelopes is reported (i.e. RC=17.3; I1=13.7; I2=10.5; C=8.4). Error bars represent the standard deviations from three independent experiments. Data from [149].

protein:ligand complexes were considered (Fig.6.18), yielding more reliable information on the conformational changes induced by ligand binding but was not exactly comparable to the blind molecular

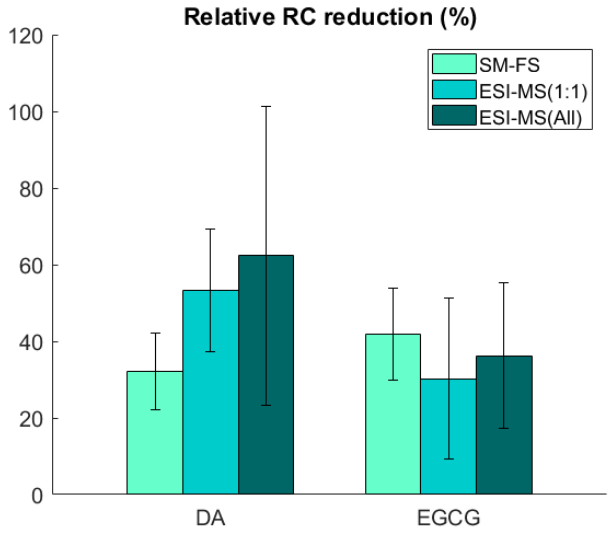

Figure 6.20.: RC reduction in response to ligand binding, relative to the free protein, as obtained by SMFS and native MS, considering the 1:1 protein:ligand complexes (1:1) or the cumulative MS data (all). Error bars represent the propagated standard deviation. Data from [149]. Figure adapted from [1].

selection performed by SMFS. Thus, *cumulative* MS data, derived by Gaussian fitting of the artificial CSD obtained by the summation of the species-specific CSDs corresponding to the different binding stoichiometries, including the free protein were considered (Fig.6.19).

In either way, the aggregated data for the unstructured (RC) component, represented as relative change from the reference condition of the protein in the absence of ligands, indicated a remarkable loss of the most disordered conformation induced by ligand binding in both the native-MS data set considered, in good agreement with the SMFS data (Fig.6.20).

For comparison to bulk spectroscopic methods, far-UV circular dichroism (CD) and Fourier-transform infra-red spectroscopy (FTIR) analyses were performed, as well. The aim of these measurements was to assess if the increment of a more compact structure (for native MS) or for a mechanically stable AS could be linked to an increase of the secondary structure. Representative results are reported in Fig.6.21.

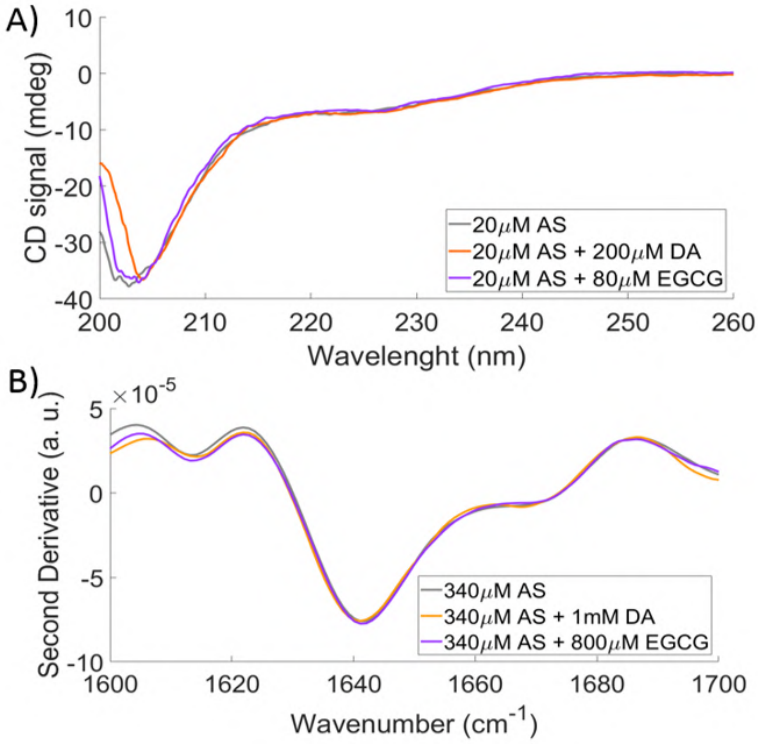

Figure 6.21.: Secondary-structure content as obtained by CD and FTIR techniques: A) CD spectra of 20 µM AS in PBS buffer in the absence of ligands (gray), in the presence of 200 µM DA (orange) or 80 µM EGCG (purple). B) Second derivatives in the Amide I region of the FTIR absorption spectra of 340 µM AS in deuterated PBS buffer in the absence of ligands (gray) and in the presence of 1 mM DA (orange) or 800 µM EGCG (purple). Figure taken from [1].

It can be noted that AS spectra in the presence or absence of the ligands, acquired by either technique under the same conditions employed in SMFS experiments, were almost superimposable. Thus, bulk methods probing secondary structure failed to capture the conformational changes induced by ligand binding in monomeric AS in solution.

7

DISCUSSION: HETEROGENEOUS ENSEMBLES OF AS PERTURBED BY LIGANDS AND MUTANTS

7.1 WT AS AND LIGANDS

The results obtained by SMFS for WT-AS provided direct evidence of the presence of different conformers of AS in solution. The structural heterogeneity of free AS in solution captured by SMFS was fully consistent with previous results by using the same approach [88, 89], as well as with results from native MS [160, 149], computational simulations [224, 225] and chemical crosslinking [174], indicating the presence of at least three different conformational states characterized by different degrees of intramolecular interactions. In analogy with SMFS, the CSD analysis of nano-ESI-MS spectra identified, in addition to the unstructured state, the presence of three different more compact components: two intermediate species (I_1 and I_2) and one compact conformation (C).

SMFS was applied here for the first time to probe the effects of the two fibrillation inhibitors DA and EGCG on AS conformational properties. The SMFS measurements for WT-AS in the presence of DA and EGCG showed a clear structural effects of AS molecule elicited by ligand binding, resulted in a rearrangement of the conformational ensemble, according to other studies for IDPs:ligands bindings [223].

Furthermore, both SMFS and native MS technique clearly indicated a loss of the most disordered component in response to ligand binding, resulting in the accumulation of the more structured (or compact, as in the case of native MS data) species. Indeed, this increase in more compact (I1, I2 and C) and more stable structure (WI and SI) could not be simply seen as the distinct contributions of van-der-Waals interactions and ordered secondary structure, differently to the proposal of *Sandal et al.* [88, 89]. Indeed, an increase of almost 30% in the SI component, as observed here in AS:DA and AS:EGCG binding, would have been detected by CD and FTIR spectroscopies, if ascribable to the formation of secondary-structure rich component. It was conceivable that the AS conformational components detected by SMFS under these conditions differed by contact order, type and number of interactions, within a picture of similar secondary-structure content. Nonetheless, the structural intermediates detected by SMFS (WI) and

native MS (I1 and I2) could not be related in a straightforward way. Similarly, it could not be proved the unambiguous identification of the SI state with the most compact structure detected in native MS (C). These could be due to the fact that the two techniques detected different physical properties: while SMFS discriminated protein structures according to their mechanical stability under an external tension (quantified by the unfolding force), native MS was affected by structural compactness (quantified by the acquired net charge). Different compaction levels could correspond to similar unfolding force and *vice versa*. Accordingly, the WI state, as detected by SMFS, was characterized by a number of variable peaks ranging from 1 to 3 different species, which could be compatible with different AS compaction states. Furthermore, the SMFS instrumental noise, related to the minimum measurable force (around 20pN) limited the minimal detectable unfolding force, below which the less stable AS compact states are counted as RC molecules, causing a possible loss of some conformations mechanically stable at low forces.

Based on these results and on the comparison of SMFS and native MS techniques, a new structural interpretation of SMFS data was proposed [1], in particular regarding the SI structure, which was not related to the presence of secondary structure. Nevertheless, the WI population detected by SMFS and the intermediate species detected by native MS do not necessarily coincide. Actually, two intermediates were detected by MS and only one by SMFS. In addition, the WI species found by SMFS did not respond to ligands, with a not substantial change in WI population, while the MS-detected intermediates did. These results indicate that SMFS was able to capture the decrease in structural disorder induced by the ligands as other technique reported in litterature [133, 143, 145], but they describe the partially structured species of the conformational ensemble in different ways. In particular, it seems that the collapsed and partially structured species detected by MS contribute cumulatively to the SI component by SMFS, while the WI component by SMFS does not find correspondence in the MS spectra. These interactions could be too weak to survive the ionization/desolvation step required for MS data.

Finally, the conformations with lower unfolding force, as detected by SMFS, could include some components with higher charge-state detected by ESI-MS. This hypothesis could be verified by comparing the two techniques in terms of the response of the RC component to the binding of the ligand. Indeed, upon binding of either ligand there was a compatible trend of loss in such a component, as observed by both techniques, in favour of more compact structures (native MS) or stronger interactions (SMFS). Indeed, it was proposed that AS changed its structure in response to DA, or most likely to by-products of DA oxidation, adopting a conformation where its N- and

C termini become closer together, thus, inhibiting AS fibrils formation [149]. Since SMFS results showed that AS was more likely to assume a compact and more mechanically stable structure upon binding to DA, it seems to be safe to suppose that this conformational change of AS may inhibit the fibril formation. A similar mechanism might drive EGCG molecules to bind to unfolded polypeptide chains and form unstructured, nontoxic oligomers, preventing the formation of ordered βsheet-rich pathological aggregates.

Based on this study, it was possible to conclude that SMFS indicated the simultaneous presence of collapsed and partially structured conformers of AS monomer in solution, consistently with other techniques, and, most importantly, revealed induced-folding transitions elicited by ligand binding. Furthermore, we showed that single-molecule protein unfolding was able to capture changes in AS conformational landscape, induced by variable solution conditions, with remarkable sensitivity and reproducibility.

7.2 AS MUTANTS

Despite the low statistic, the preliminary data of SMFS of pointy mutated AS allowed us to infer that our experimental design was able also to describe the conformational rearrangements of the AS with pathological point mutations.

7.2.1 *A30P and A53T*

The distributions of the population of A30P and A53T AS mutants in our investigation were substantially identical (~ 42% RC, ~ 16% SI and ~ 42% WI).

The effect of the point mutations in the N-terminal region (A30P and A53T) on the AS behaviour were investigated in the past by using a construct very similar to the one employed in this work [88, 89] and a more elegant polyprotein construct in which the AS was mechanically protected by a mechanostable carrier [169]. In [88] only the A30P mutant was evaluated, showing an increase of the SI population (from 8% for WT-AS to 37.5% for the A30P-AS) at the expense of the WI conformation, while our results showed an increase of both SI and WI populations. One year later, the same group [89] repeated the analysis including the A53P mutant, depicting a quite different scenario: for the A30P mutant was found a slight increase in RC (from 41% to 53%), an important increase in SI (from 6% to 28%) and a decrease in WI (from 53% to 19%) while for the A53P mutant the RC conformation was essentially constant, the WI decreased to the final valure of 21% and the SI increased to 36%. However it should be noticed that the buffer conditions were different from the one used in this

work, 10mM Tris-HCl pH 7.5 and the AS protein is very susceptible to environmental changes.

Furthermore, the group of Carrión-Vásquez [169] obtain very similar results for both the two mutants (40% NM and 60% M, in PBS pH 7.4), showing an important decrease in the non-mechanically stable conformations. If we consider the WI and SI as M population we obtained a percentage (42%+16%=58%) very closed to the one accepted in literature.

7.2.2 *E83A*

The third mutant, the E83A, recently discovered, showed a quite different behaviour with respect to other mutants: the principal feature of E83A-SMFS was a decrease in RC followed by an increase in SI populations, while the WI was constant within the error. It was the first SMFS study of this mutant. The difference with the other mutants could be imputed to the its position in the NAC domain of AS with respect to A30P and A53T mutations which are located in the N-terminus. Indeed, different studies [221, 222] showed that the presence of mutations in the N-terminal portion of the AS accelerated oligomerization, while the E83A mutant seemed to increase the tendency to form β-structures involved in both fibril formation and aggregation. From this point of view it could be proposed that the AS oligomerisation might be driven by an increasing of both WI and SI, while the fibrillation might be principally imputed to the presence of AS in the SI state.

7.3 CONCLUSIONS AND FUTURE PERSPECTIVE

Single-molecule description of AS conformational ensemble in solution detected differently structured components, consistent with the different degrees of compactness suggested by native MS. Our results might provide valuable constraints for computational simulations of IDP conformational ensembles in the presence or absence of interactors and new insight into the PD development linked to gene mutations.

As a perspective, knowing that AS:DA interactions are stabilized by long-range electrostatic interactions with glutamate 83 (E83) in the NAC region, it would be of particular interest to study the effects of DA on the E83A mutant. AS-mutant data suffered from a quite low statistic, which should be increased. Moreover, we will test a forth mutant, the E43K and we produced a biotinilated polyprotein construct which should be suitable to be used under the MT setup. This MT-based SMFS should provide information on the AS structure at lower forces, as well as kinetics and cyclic experiment of folding-refolding.

Part III

NANOMECHANICAL CHARACTERISATION OF DIAMINOPURINE-SUBSTITUTED DNA

8

DIAMINOPURINE-SUBSTITUTED DNA

In this part of the Thesis I report the main results of the nanomechanical characterisation of a Diaminopurine-substituted DNA project, which have been published on Biophysical Journal [2].

In this work a systematic comparison between a wild-type DNA and a diaminopurine substituted (DAP) DNA was performed, in terms of both thermal stability and nanomechanical properties, measured at almost-null, low and high forces by means of both AFM and MT instruments.

The exotic 2.6-diaminopurine nucleobase (DAP) is an adenine-analogue which, when incorporated into double strand DNA (ds-DNA), forms a triple hydrogen bond with thymine (Fig.8.1). This non canonical DNA is an interesting substrate both biologically (in the cyanophage S-2L DAP substitution naturally occurs) and structurally for nanoscale engineering (i.e. interactions between DNA and proteins, drug candidates or even as novel dopants in DNA-based nanoelectronics). Altough this modification, which leads a full three hydrogen bonds all along the DNA double helix, could also change the dsDNA mechanical conformation and stability without losing the sequence specificity effect.

Using a combination of single molecule tools, i.e. atomic force microscopy (AFM) and magnetic tweezers (MT) nanomechanical assays, and standard bulk approach as circular dichroism spectroscopy and thermal denaturation we demonstrated that DAP substitution increases the flexural rigidity of dsDNA also facilitates conformational shifts, which manifest as changes in molecule length. The decrease in bending flexibility results in DAP-containing molecules behaving as stiffer entropic springs, characterized by longer persistence lengths, and forming larger plectonemic loops upon supercoiling. In the almost-static case (AFM), in which no external force is applied to the molecule, the contour length of DAP-DNA appears shorter than wild-type (WT)-DNA; while under low tension tension (Forces below 2 pN), they also have similar dynamic contour lengths. At tensions above 60 pN, WT-DNA undergoes characteristic overstretching because of severe strand denaturation and spontaneous adoption of a conformation S-DNA. Cyclic overstretching and relaxation of WT-DNA at near-zero loading rates typically yields hysteresis, indica-

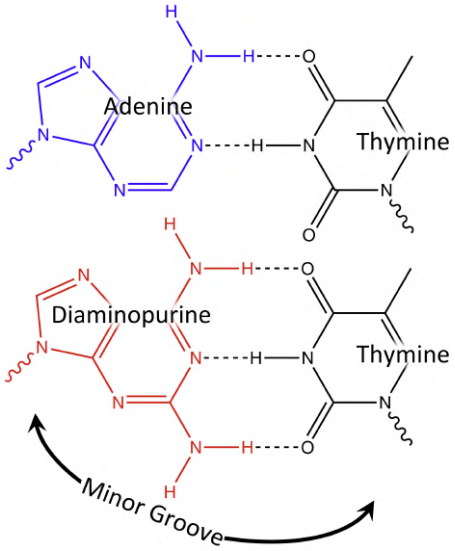

Figure 8.1.: 2,6-diaminopurine (bottom) is an analog of adenine with an additional amino group at position 2 of the purine molecule. When paired with thymine, it forms an additional hydrogen bond along the minor-groove side of the molecule. Fig. taken from [2].

tive of tension-induced melting; conversely, cyclic stretching of DAP-DNA showed little or no hysteresis, consistent with the adoption of the S-form, similar to what has been reported for GC-rich sequences. In physiological salt conditions, evenly mixed AT/GC DNA typically overstretches around 60 pN. GC-rich sequences overstretch at similar if not slightly higher tensions. In this work we assessed that DAP-DNA overstretches with a force of 52 pN. In summary, DAP substitution decreases the overall stability of the B-form double helix, biasing toward non-B-form DNA helix conformations at zero tension and facilitating the B-to-S transition at high tension.

9

MATERIALS AND METHODS: DIAMINOPURINE-SUBSTITUTED DNA

9.1 DNA PREPARATION

Normal and DAP-substituted DNA sequences were produced using a standard polymerase chain reaction (PCR). DAP-substituted DNA was produced by replacing dATP with 2,6-diaminopurine-5'-triphosphate (Trilink Biotechnologies, San Diego, CA) at the same concentration (200μM) as the other three dNTPs (Fermentas, Waltham, MA).

9.2 MELTING TEMPERATURE CHARACTERIZATION

The melting temperature of wild-type and DAP substituted sequences was determined from fluorescently detected melting curves [226, 227]. Three 155, 147, and 156 base-pair-long portions of the pBR322 plasmid were studied, containing 40%, 54%, and 65% GC-content, respectively. The sequences were amplified using Taq DNA Polymerase (New England Biolabs, Ipswitch, MA) in ThermoPol buffer. Following PCR amplification, DNA samples were purified with a Qia Quick PCR cleanup kit (Qiagen, Germantown, MD) and eluted in 10 mM Tris-HCl pH 8.5. Amplicon lengths were verified by gel electrophoresis and concentrations measured by UV absorption using a Nanodrop Lite (Thermo Scientific, Waltham, MA). DNA was diluted to $10 ng \mu L^{-1}$ in 10 mM Tris-HCl pH 7.4, with 15 mM KCl. Two intercalating dyes were used to assess melting: Syto-84 (useda a 1μM) and Sybr-Green I (used over a range of 0.0005 - 0.01% v/v 10,000x)(Thermo Scientific). For each sequence and type at least 6 samples were measured and averaged. Fluorescent intensity was recorded using a BioRad C1000 qPCR machine (BioRad, Hercules, CA) over a temperature range of 60-95°C in 0.5°C increments.

9.3 DNA FOR MAGNETIC TWEEZER AND AFM EXPERIMENTS

Magnetic Tweezer (MT) and Atomic Force Microscopy (AFM) experiments measuring DNA mechanical parameters and overstretching

were performed using 4642 bp-long (hereafter 4.6 kb) DNA fragments. For the MT experiments, tethers were constructed from three components: a 4.6 kb long core fragment containing either wild-type or DAP-DNA, and two ~ 1kb flanking tails containing biotin-or digoxigenin-11-dUTPs. The core fragment was prepared by PCR with Long Amp™(New England Biolabs) using the pKLJ12wt plasmid and primers 5′ -AGCGTTGGCGCCGATTGCAGAATGAATTT and 5′ TGGGATCGGCCGAAAGGGCAGATTGATAGG, which contain *KasI* and *EagI* restriction sites (underlined), respectively. A single major amplicon around 4.6 kbp was produced (Fig. 9.1). The biotin and digoxigenin-labeled tail fragments were also produced by PCR using Taq polymerase in standard buffer (NEB). PCR solutions were supplemented with biotin-11-dUTP (Fermentas) and digoxigenin-11-dUTP (Roche, Indiannapolis, IN) in a 1:9 ratio with respect to dTTP. The biotin-labeled fragment was amplified from pUC19 using the primer pair 5′-ATGATCCCCCATGTTGTGCA and 5′-TCAAGACGATAGTTACCGGATAAG to create a 1.8 kb biotin-labeled amplicon with a central KasI site. The digoxigenin-labeled fragment was amplified from pBluKSP using the primer pair 5′-TGGGTGAGCAAAAACAGGAAGGCA and 5′-GCGTAATCTGCTGCTTGCAA to create a 2 kb dig-labeled amplicon with a central EagI site. Thermocycle conditions are listed in Table S4.

Following PCR amplification and column purification (Qia Quick PCR cleanup; Qiagen) the core and tail fragments were digested with KasI and EagI-HF restriction enzymes (New England Biolabs), purified again, and concentrations were measured by UV absorption. Restriction of the tails yields roughly 1 kb fragments with a single KasI or EagI sticky-end. Restriction of the core fragment produces a 4.6 kb sequences with KasI and EagI sticky-ends at opposite ends. Note that this fragment sets the effective tether length in MT experiments, because the randomly bio- or dig-labeled handles will almost completely attach to the magnetic bead and the chamber surface, marginally contributing to the DNA extension. This is confirmed by DNA extension MT measurements (as seen in Fig.10.5). As consequence, we expect that the MT measurements are mainly determined by the core fragments.

To assemble the final tether, approximately 600 ng of the core fragment was mixed with 250 ng of each tail along with 30 U of T4 DNA ligase (NEB) and the recommended buffer to a total volume of 30 µL and incubated at 16°C for 4 hours. After heat-inactivation at 65°C for 20 min, the ligation mixture was used directly in the micro flow-chambers.

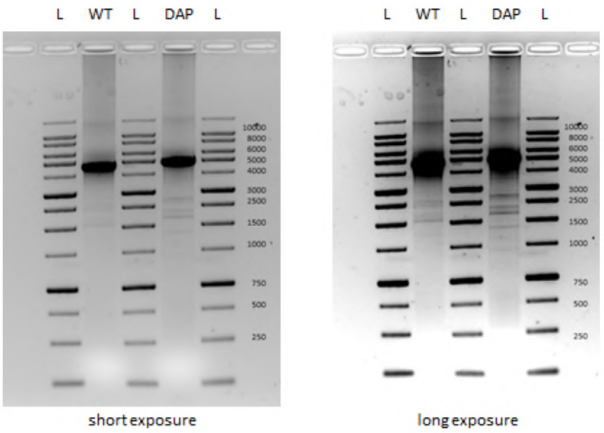

Figure 9.1.: A short exposure (left) of a 1% agarose gel of the products from PCR reactions with dCTP, dGTP, dTTP, and either dATP (WT) or dDAPTP (DAP) reveals a major amplicon near 4.5-5 kbp and other minor amplicons which are more abundant in the DAP sample. Short amplicons are also visible in the AFM images (Fig.10.2, panel B). A longer exposure (right) of the same gel displays a saturated signal for the desired amplicon while other bands are still quite faint. This indicates the relative homogeneity of the desired amplicon.

9.4 MAGNETIC TWEEZING

The Magnetic Tweezer (MT) used in this work was built using an inverted, infinity-corrected microscope. A set of permanent magnets was mounted above the sample, between the stage and condenser optics. These magnets could be translated along and rotated around the optical axis to control the tension and supercoiling of the DNA tethers [228, 229]. The flow-chamber used for the MT experiments consists of a 50 µL capillary functionalized with anti-digoxigenin antibody (Roche, Indianapolis, IN) [228]. DNA-tethers were formed by incubating the biotin and digoxigenin end-labeled fragments with streptavidin-coated superparamagnetic beads (Invitrogen, Carlsbad, CA) for ~ 15 min. Conjugated tethers were isolated with a magnet and re-suspended in 500 µL PTE buffer (phosphate-buffered saline pH 7.4 with 1mM EDTA and 0.1% Tween). For low force experiments 1 µm diameter magnetic beads were used, while for overstretching measurements (where forces in excess of 50 pN were needed) 10µm beads (Invitrogen) were used. Finally, tether suspensions were incubated in the capillaries for about 1h before flushing out unbound tethers with 8-10 mL PTE buffer. All MT measurements were carried out in PTE buffer.

9.5 AFM BASED IMAGING

Nano-scale images of wild-type and DAP-DNA deposited onto mica surfaces were acquired using a Nanowizard II (JPK Instruments, Berlin, Section 1.2). WT and DAP molecules were diluted in 5 mM $MgCl_2$ in MilliQ water to a final concentration between 0.2-0.3 $ng\mu L^{-1}$. 10 μL of each solution was deposited onto freshly cleaved mica substrates and incubated at room temperature for 5-10 min. The samples were then rinsed with 0.22 μm-filtered deionized ultra-pure water and dried with a gentile nitrogen flow.

Images were acquired with the AFM operating in tapping mode, in air, using stiff silicon cantilevers (RTESO-Veeco, resonant frequencies of 300-350 kHz and spring constant $\sim 40 N \cdot m$). 8 x 8 μm^2 images were collected using a 0.5-1 Hz scan rate with 2048 x 2048 pixel resolution. Several hundred DNA molecules were recorded from several widely-spaced areas of each sample and processed with a standard flatten filter in the JPK Data Processing software.

To morphologically characterize the two samples, we traced the DNA contours using a custom-tracing routine [230] and calculated the contour length, L_0 [231]. Only molecules in the range 1.1-1.8 μm surrounding the expected length were included. The average end-to-end distance of segments of the DNA molecules as a function of their separation along the contour was analyzed to estimate the persistence length (L_P) according to the worm-like chain model (Appendix D.3).

9.6 CIRCULAR DICHROISM SPECTROSCOPY

Wild-type and DAP DNA molecules produced for the melting experiments and the 4.6 kb core fragments from the AFM/MT experiments were diluted to concentrations ranging from 10-30 ng μL^2 in 10 mM Tris-HCl pH 7.4 with 150 mM KCl. The CD spectra of the respective samples were measured using a Jasco J-810 spectrometer (Jasco Corporation), in a quartz cuvette and blanked against the spectrum of Tris-KCl buffer without DNA. The spectrum of each sample was measured three times at wavelengths 210-310 nm (in 0.2 nm increments), averaged and smoothed with a Savitzky-Golay filter of third-order using a 31 step (6 nm) window width.

10

RESULTS: NANOMECHANICAL CHARACHTERISATION OF DNA-ANALOGUE

10.1 THERMAL DENATURATION

Under the typical DNA B-form structure Watson-Crick base pairing, the two backbones of a DNA molecule are held together by hydrogen bonds formed between the opposites base pairs. As a results of this model, it is reasonable to expect that the melting temperature may be proportional to the total base-pairs hydrogen bond in the molecule under study, which, for a standard DNA molecule reflects the number of A-T and G-C base pairs. However, it has been demonstrated that also the stacking interactions between the A-T and G-C duplexes play a role in changing the melting temperature, arising a sequence-dependent effect [232, 233]. Consequently, the DAP-substitution into a DNA molecule, even relatively short, could be expected to affect melting temperatures at least in three different ways: the 3 hydrogen bond for all the base-pairing stabilized the molecule, the stacking interactions of adenine and DAP are likely different, and the presence of an extra amino group may alter the conformational stability of B-DNA.

The melting temperature (T_m) of three sequences containing 40%, 54%, and 65% GC were compared, as shown in Fig.10.1 where measured (red and blue dots) and calculated (yellow and purple dots) values of T_m are presented as a function of GC percentage. The fraction of helical versus single stranded DNA was measured by fluorescence in the presence of the intercalating dye Syto-84, which has been shown to shift melting temperatures by less than 0.6°C at 1μM.

For relatively short sequences, as the ones which are used here, the melting process is approximately binary and the full sequence separate at the same time. With that approximation, melting temperatures can be estimated by

$$T_M = \frac{\Delta H}{\Delta S + R \ln \frac{[dsDNA]}{2}} - 273.15, \tag{10.1}$$

where T_m is the temperature at which 50% of dsDNA molecules will be melted, ΔH is the sequence dependent enthalpy, ΔS is the sequence

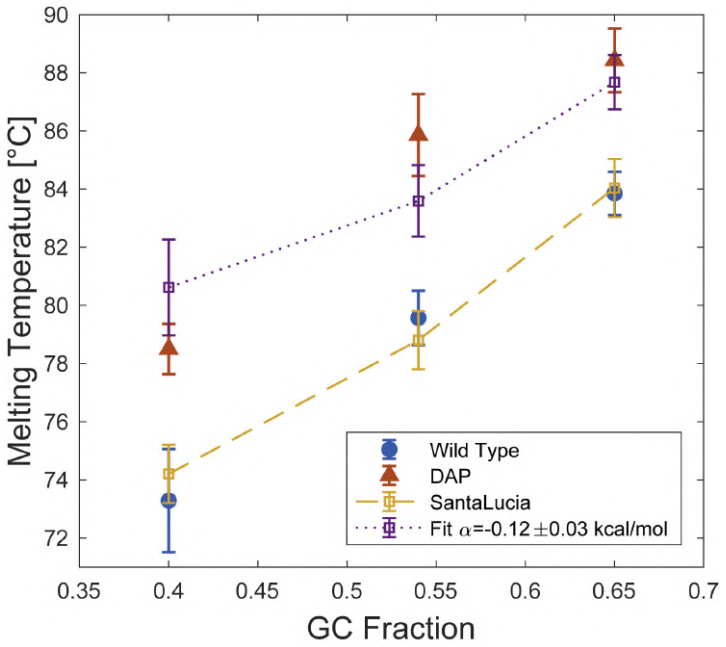

Figure 10.1.: Melting curves of three sequences of WT (blue, solid circle) and DAP-DNA (red, solid triangles) with GC fractions of 0.4, 0.54, and 0.65 were measured using the fluorescence of Syto-84 (1 μM); error bars indicate SD. WT results closely match T_m predicted by the SantaLucia nearest-neighbor model (yellow); error bars indicate model accuracy. The DAP melting temperatures were fit with a modified SantaLucia model (purple). A least-squares fit yields $\alpha = 0.12 \pm 0.03$ kcal $(\text{mol})^{-1}$. Data obtained by Dr. Daniel Kovari in Dunlap Lab, Emory University, Atlanta. Fig taken from [2].

dependent entropy, R is the gas constant and [dsDNA] is the concentration of dsDNA in solution.

Under the SantaLucia nearest-neighbour model, ΔH and ΔS are equal to the sum of empirically determine nearest-neighbour terms and are corrected to account for salt-dependent effects [234, 235]. This model yields a predictions of the three sequences melting temperature within $\pm 0.8^{o}C$ of (Fig.10.1 blue and yellow dots).

DAP-substitution increases melting temperatures without eliminating sequence-dependent effects, increasing the T_m of each sequence by 6-8°C. This clearly indicate that the melting temperature is not only dependent to the number of H-bonds but it is strongly sequence-dependent.

To underscore this point a modified version of the SantaLucia model has been used. Under this extended model, the enthalpy of each A-T pair is supplemented with an additional factor (α) accounting for both the effect of the additional hydrogen bond, changes in stacking energy, and perturbations to conformational stability.

A least-squares fit to the DAP melting data yields $\alpha = -0.12 \pm 0.03 kcal(mol)^{-1}$. While even this model does not perfectly predicts the T_m for each sequence, it does preserve the general trend.

10.2 ZERO-FORCE REGIME: ATOMIC FORCE MICROSCOPY

A zero-force characterisation has been performed by means of AFM imaging. WT and DAP-DNA molecules were deposited onto freshly cleaved mica surfaces and imaged in tapping mode, obtaining both the static contour (L_c) and persistence lengths (L_p). Representative images of both WT and DAP samples are shown in Fig.10.2. Both the L_c and L_p were measured by tracing each DNA filament using a MATLAB custom-written image recognition routine and computing the integrated length [230].

For the L_c calculation, molecules that were significantly longer or shorter than the expected range were discarded from the data set. An histogram of contour lengths for WT and DAP molecules is included in Fig.10.3A. As indicated, DAP-substitution decreased the average contour length from $1.45 \pm 0.12 \mu m$ to $1.30 \pm 0.12 \mu m$ ($p = 3.07 \cdot 10^{-4}$), corresponding to axial rises per base pair of $3.11 Å(bp)^{-1}$ and $2.79 Å(bp)^{-1}$ respectively, consistent with previous measurements [236].

The persistence length $L_p$was estimated from the measured mean square displacement between points along the contours of molecules in two-dimensions [230]. For a worm-like chain confined to two dimensions, the average squared distance $\langle R_L^2 \rangle$ between two points along the chain is given by

$$\langle R_L^2 \rangle = 8 L_p^2 \left(\frac{L}{2L_p} - 1 + e^{-\frac{L}{2L_p}} \right) \tag{10.2}$$

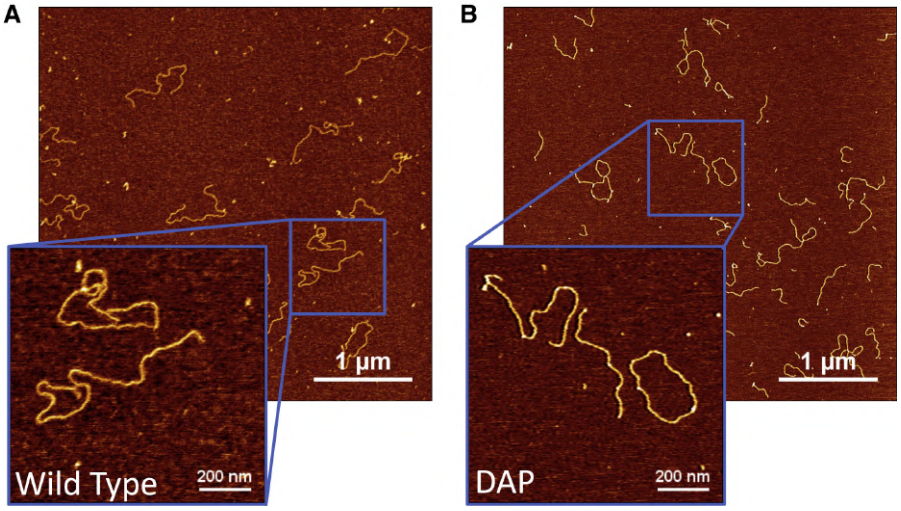

Figure 10.2.: Representative images of (A) wild-type and (B) DAP-DNA molecules captured via AFM. Fig. taken from [2].

where L is the curvilinear distance along the chain between the points, and L_p is the molecule's persistence length [237, 238]. Using the x-y coordinates generated by the tracing routine, it is straightforward to compute the distance between points separated by increasing contour lengths. The data for molecules from either WT or DAP samples were pooled, and the means of the squared end-to-end distances were computed as a function of separation along the contour length and fit to Eq.10.2 (results shown in Fig.10.3B). Excluded volume effects were negligible given the low DNA concentration and the limiting of contour length separations to less than 200 nm when fitting Eq.10.2 [239]. WT-DNA yielded $L_p = 56.2 \pm 0.1 nm$, while DAP-substitution increased the persistence length to $L_p = 79.9 \pm 0.3 nm$ (95% confidence interval). Analysing overlapping or non-overlapping segment data produced similar estimates (Fig.10.4 and Table 10.1).

10.3 LOW-FORCE REGIME: MAGNETIC TWEEZERS

The mechanically characterization of DAP sobstituited-DNA has been performed via Magnetic Tweezers (MT)-based force spectroscopy. 4.6 kbp DNA tethers were subjected to forces ranging over 0.001-5 pN and their end-to-end lengths were measured. Only molecules in the range 1.1-1.8 µm, surrounding the expected 1.5 µm length, were included in the analysis. Below 10 pN DNA acts as an entropic spring and is well characterized by a worm-like-chain (WLC) polymer model (see Appendix D for details). The force-length dependence for a WLC is approximately

$$F(L_e) = \frac{k_B T}{L_p}\left(\frac{L_e}{L_0} + \frac{1}{4}\left(1 - \frac{L_e}{L_0}\right)^{-2} - \frac{1}{4}\right) \tag{10.3}$$

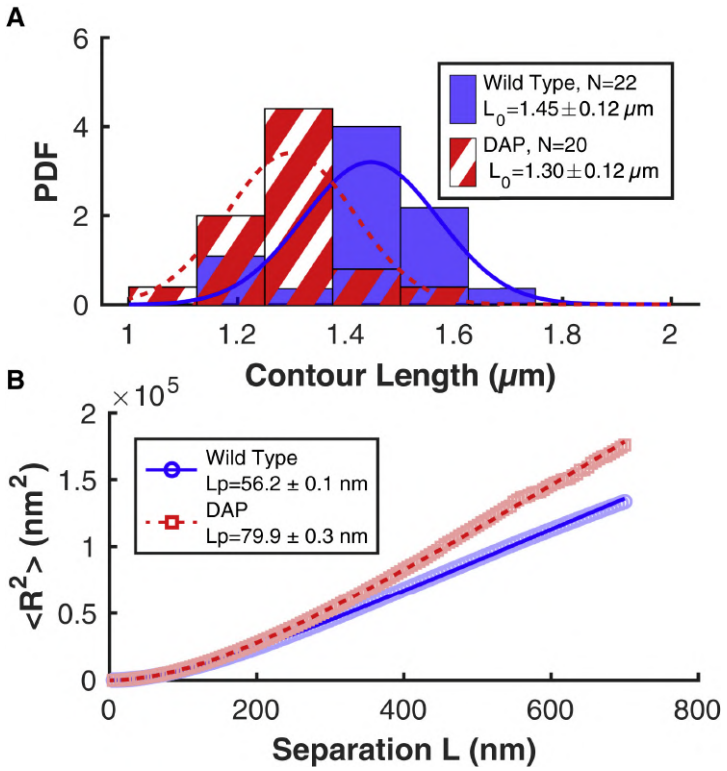

Figure 10.3.: Contour and persistence length measured via analysis of AFM images. (A) Contour length was calculated by tracing individual molecules in each AFM image. DAP substitution reduced the length. The distributions are statistically different in a two-tailed t-test ($p = 3.07 \cdot 10^{-4}$). (B) Mean-squared distance versus separation along each molecular contour is shown. Data points (blue circles and red squares) correspond to the average square distance, aggregated for all molecules of a given type (WT and DAP, respectively). Data was fit to Eq.10.2, yielding estimates of the average persistence length; ± error indicates 95% confidence interval. Fig. taken from [2].

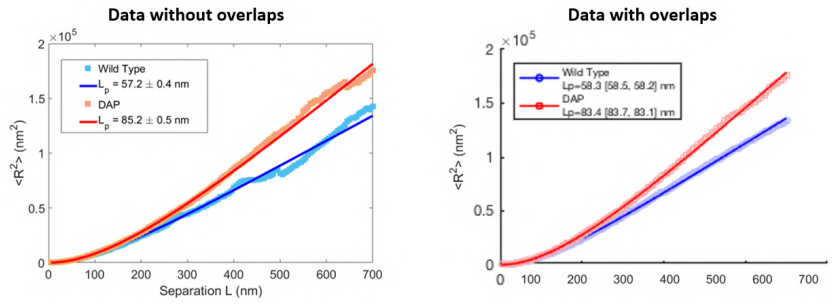

Figure 10.4.: Plots of the average, squared end-to-end distance as a function of the contour length of segments within DNA molecules imaged using the AFM. Selecting segments with no overlaps (left) produced more variation in the curves with respect to selecting all available segments (right). However, the persistence length values obtained by fitting the curves were quite similar (Table 10.1). Fig. taken from [2].

	Data with Overlapping		Data without Overlapping	
Max L separation	WT	DAP	WT	DAP
nm	nm	nm	nm	nm
100	56.0 ± 0.8	80.0 ± 0.5	55.6 ± 0.4	79.8 ± 0.2
200	56.4 ± 0.5	80.2 ± 0.3	56.2 ± 0.1	79.9 ± 0.3
300	58.4 ± 0.6	83.3 ± 0.4	57.9 ± 0.2	80.7 ± 0.3
400	57.2 ± 0.4	85.2 ± 0.5	58.3 ± 0.1	83.4 ± 0.3

Table 10.1.: Fitting parameter L_c and L_p for both overlapping and not overlapping segments . Parameters were obtained of the average, squared end-to-end distance of segments within DNA molecules imaged using the AFM as a function of the contour length of those segments were fit with two-dimensional worm-like chain models limiting the maximum contour length as indicated for each row. Limiting the maximum separation value between 100-400 nm only slightly changed the fitted persistence length values.

where L_e is the end-to-end distance, L_0 is the molecule's contour length, and L_p is the persistence length. For each molecule, force versus length measurements were fit to Eq. 10.3 yielding estimates of both L_0 and L_p. Representative measurements and accompanying fits for a WT and DAP molecule are included in Fig.10.5. DAP-substitution increased persistence length from 45 ± 4 nm for WT to 56 ± 5 nm for DAP ($p = 5.22 \cdot 10^{-16}$). However, dynamic contour length only decreased from 1.51 ± 0.13 μm to 1.46 ± 0.12 μm, a change that is not statistically distinct ($p = 0.103$, see Fig.10.5B, C).

Both MT and AFM reveal a significant increase in the persistence length values after DAP-substitution, indicating that DAP-DNA molecules are stiffer (i.e. harder to bend). This result is consistent with what has been reported previously for experiments using AFM and optical tweezing [236, 241]. Similarly, previously reported MT experiments measuring supercoiled-DNA found that DAP-substitution increased the size of plectonemic writhes, which is consistent with DAP-DNA being harder to bend [242].

10.4 HIGH FORCE REGIME: MAGNETIC TWEEZERS

The DNA overstretching transition is characterized by a roughly 1.7x increase in the molecule length when the applied force exceeds a critical value. Under near-physiological conditions, overstretching typically occurs around 60-70 pN for WT-DNA. This transition exhibits both force-induced melting, wherein the phosphate backbones separate enabling each strand to stretch independently, as well as the formation of a double-stranded form called S-DNA. The propensity of a DNA molecule to form S-DNA rather than undergo force-induced melting is sequence dependent. Indeed, simulations have shown

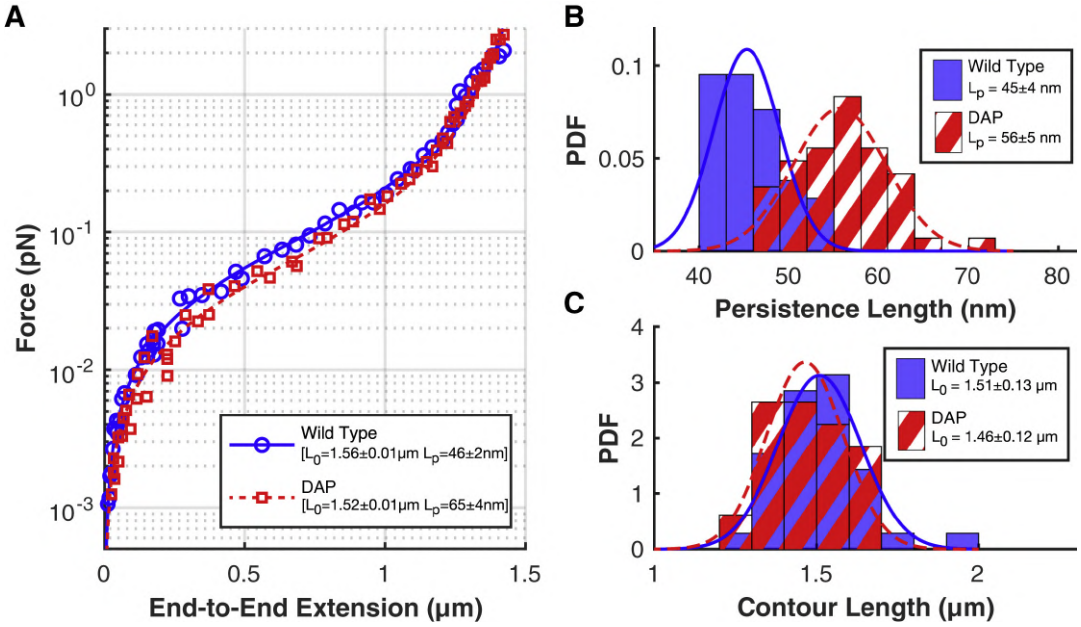

Figure 10.5.: WLC fits of force versus length MT data. (A) Representative force versus length data for wild-type (blue) and DAP (red) DNA molecules with corresponding WLC fits (solid line for WT and dashed line for DAP) are shown. Fit errors correspond to 95% confidence limits. (B) Histogram of fit persistence length values is shown; DAP substitution increases Lp to 56 ± 5 from 45 ± 4 nm for WT. Distributions are statistically distinct, $p = 5.22 \cdot 10^{-16}$. (C) Histogram of fit contour length values . DAP substitution does not significantly decrease dynamic contour length, p = 0.103. N = 35 for WT and N = 48 for DAP in both histogram plots. Figure taken from [2].

that while G-C pairs persist through the transition, A-T pairs often form melted bubbles [243]. Moreover, cyclic pulling experiments demonstrate that force-induced melting is associated with hysteresis, such that once melted, the complementary strands re-anneal slowly [246, 245, 244].

To investigate how DAP affects overstretching, the force-extension behaviour have been analyzed in the high-force regime both for WT and DAP DNA. Torsionally constrained molecules (verified by a twist versus extension curve) were excluded from the analysis because torsional constraint fixes the helical linking number, forcing hyper-extended molecules to adopt a so-called P-DNA configuration [247]. Nicks occur randomly during the preparation of the DNA tether and it has been estimated that less than 26% of the molecules analyzed had more than one nick. Thus, nicks did not strongly bias the overstretching measurements nor the observed hysteresis. Remarkably, DAP-DNA overstretched at a significantly lower transition force than WT-DNA. The overstretching transition force (F_{OS}) was characterized by calculating the midpoint of the transition plateau. As shown in Fig.10.6A, DAP-substitution lowered the transition force to $F_{OS} = 52 \pm 3$ pN compared to $F_{OS} = 60 \pm 2$ pN for WT-DNA for data collected in 150 mM PBS.

Cyclic extension and relaxation also revealed that DAP-DNA exhibits very little hysteresis through the overstretching transition. The lack of hysteresis in the DAP curves indicates that peeling does not occur even in the presence of nicks, in support of DAP bases favoring S-DNA formation rather than melting. By comparison, nearly all WT molecules show significant hysteresis, indicating that force-induced melting constituted the bulk of extensional increase. To quantify the difference in DAP and WT behavior we calculated the area between the extension and relaxation traces for each molecule (see Fig.10.6C, D for illustration). The MT essentially operate as force-clamps. In these experiments, each data point is recorded over a period of 45-60 seconds, with at least 1 sec between successive measurements. The entire hysteresis measurement requires at least 20 min. Consequently, the molecules were stretched at almost zero loading-rate and the DNA remains overstretched for several minutes. Results are shown in Fig.10.6B. The smaller secondary plateau in Fig.10.6D at about 56 pN appeared in some, but not all measurements; therefore we do not speculate as to its origin. In summary, WT-DNA had an average hysteresis area of $19 \pm 9 \mathrm{pN} \cdot \mu\mathrm{m}$, while DAP-DNA averaged $2 \pm 1 \mathrm{pN} \cdot \mu\mathrm{m}$ ($p = 2.73 \cdot 10^{-5}$).

10.5 CD SPECTROSCOPY

The circular dichroism spectra of wild-type and DAP versions of the 40%, 54%, 65% GC, and 4.6kb core fragments were measured over

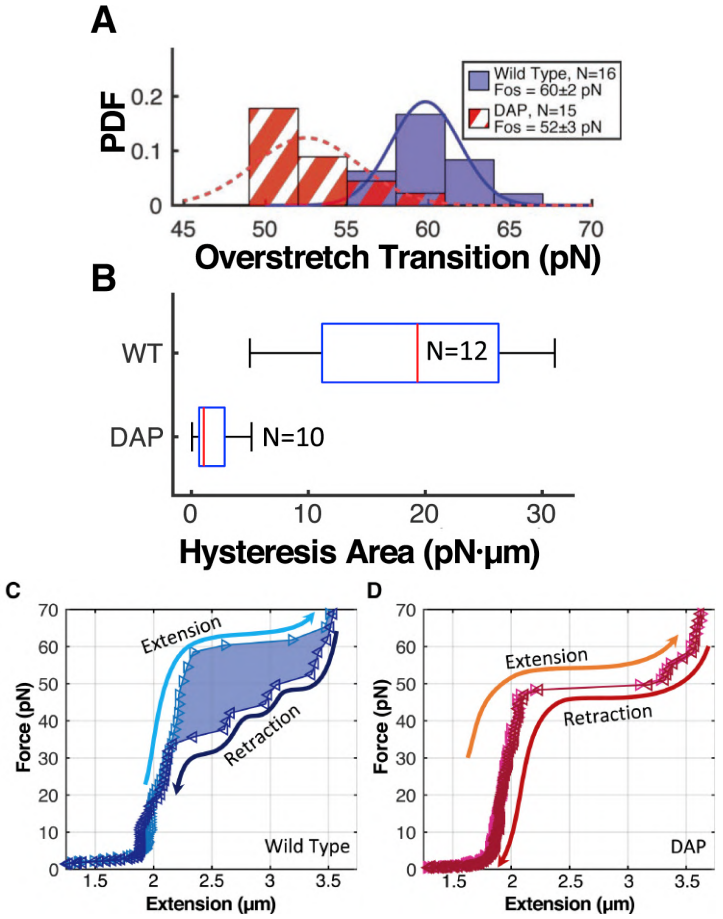

Figure 10.6.: Overstretching under high external force. (A) Histogram of overstretching transition force is shown. Distributions are statistically distinct, $p = 2.88 \cdot 10^{-8}$. (B) Whisker plot of area between extension and relaxation curves are shown. Nearly all WT molecules exhibited significant hysteresis $H = 19 \pm 9 pN \cdot \mu m$; DAP molecules showed little hysteresis $H = 2 \pm 1 pN \cdot \mu m$. (C) Example force-extension cycle for WT-DNA is shown. As tension is increased (light blue trace), the molecule undergoes force-induced melting. Upon relaxation (dark blue), molecule length decreases with a different characteristic curve owing to the presence of single-stranded DNA. (D) Example force-extension cycle for DAP-DNA is shown. Little hysteresis is observed between extension and relaxation, indicating basepair persistence and the adoption of the S-form

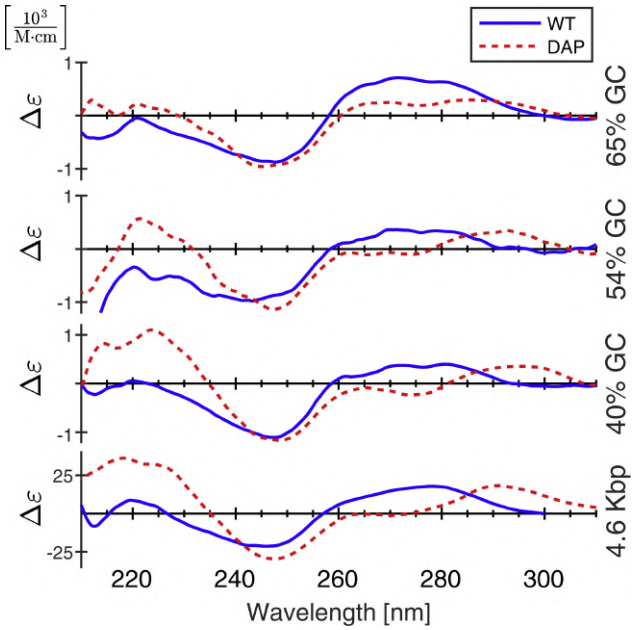

Figure 10.7.: CD spectra of WT (blue, solid lines) and DAP-substituted molecules (red, dashed lines).

the range $210 - 310$nm. Results are shown in Fig.10.7. The spectra of all four wild-type molecules show the expected behavior of B-DNA, specifically a local-maximum in the $270 - 280$nm range and a local-minimum around 250nm [248]. The spectrum of 4.6kb DAP molecules is very similar to the spectra reported by Peters *et al.* and Virstedt *et al.* for DAP-substituted molecules without extensive AT or GC repeats, namely the 4.6kb DAP-spectra shows a peak near 290 nm, a shoulder between $260 - 270$nm and a minimum near 250 nm [236, 241]. Interestingly, the 40%, 54% and 65% GC sequences show a progressive shift from non-WT behavior towards a WT-like spectrum. 40% molecules, which have the most DAP, show a peak at 295 nm, a minimum at 248 nm, and a modest local-minimum at 274 nm. As DAP content decreases (and GC-content increases) the minimum at 274 nm fades.

The measured mechanical characteristics for wild-type and DAP-substituted DNA are summarized in Table 10.2. The values reported in the text refer to mean $\pm$ standard deviation.

Parameter	Unit	Type	Mean	Std. Dev.	N	95% CI	p-value
L_0 (MT)	μm	WT	1.51	0.13	35	± 0.04	0.103
		DAP	1.46	0.12	48	± 0.03	
L_0 (AFM)	μm	WT	1.45	0.12	22	± 0.05	$3.07 \cdot 10^{-4}$
		DAP	1.30	0.12	20	± 0.05	
Axial Rise (MT)	$Å(bp)^{-1}$	WT	3.25	0.27	—	± 0.09	—
		DAP	3.16	0.27	—	± 0.07	
Axial Rise (AFM)	$Å(bp)^{-1}$	WT	3.11	0.12	—	± 0.11	—
		DAP	2.79	0.12	—	± 0.11	
L_p (MT)	nm	WT	45	4	35	± 1.2	$5.22 \cdot 10^{-16}$
		DAP	56	5	48	± 1.5	
L_p (AFM)	nm	WT	56.2	—	—	± 0.1	—
		DAP	79.9	—	—	± 0.3	
F_{OS}	pN	WT	60	2	16	± 1.1	$2.88 \cdot 10^{-8}$
		DAP	52	3	15	± 1.6	
Hysteresis Area	$pN \cdot \mu m$	WT	19	9	12	± 5.08	$2.73 \cdot 10^{-5}$
		DAP	2	1	10	± 1.07	

Table 10.2.: Summary of results

11

DISCUSSION: DAP SUBSTITUTION INDUCES DNA STRUCTURAL REARRANGEMENT

The melting experiments clearly demonstrate that DAP-substitution increases the thermal stability of dsDNAs but does not raise melting temperatures to the expected level of equivalent length poly-GC fragments. Fitting the melting data against the modified SantaLucia model shows that the DAP-substitution stablizes the per-base hybridization energy (α in Fig.10.1) by $\Delta\Delta E = -0.12 \pm 0.03 kcal(mol)^{-1}$, but it does not reach the values expected by "dangling-bond" assay [249]. The results suggest that factors such as conformational instabilities (resulting from steric clashes in the minor groove, for example) partially offset the stabilizing effects of the extra hydrogen-bond and stronger base-stacking.

In addition to the increased thermal stability, the DAP-DNA also resists force-induced melting. While WT molecules exhibited significant hysteresis when relaxing from the overstretching transition, DAP molecules show little to no hysteresis, suggesting that DAP-DNA molecule's two polynucleotide backbones remain hybridized during overstretching [250].

The low-tension AFM and MT experiments confirm that hat DAP-substitution stiffens the molecule against lateral bends and increases persistence length. However, the results at higher tension also reveal that DAP-substitution lowers the energetic barrier of conformation changes, lowering its resistance to overstretching.

The CD-spectra of WT and DAP molecules shown in Fig.10.7 confirm that DAP-substitution alters the helical conformation, although the spectra of DAP-DNA are not indicative of A-DNA. In line with what has been reported by others [249], DAP-DNA shows a peak near 290nm, a local minimum near 275nm and a minimum at 250nm, distinctly from both the B-DNA and A-DNA configurations.

Interestingly, while DAP-substitution resulted in a significant reduction in contour length as measured by AFM (from $3.11Å(bp)^{-1}$ down to $2.79Å(bp)^{-1}$, a 10% decrease), it only had a marginal effect when measured with MT ($3.25Å(bp)^{-1}$ to $3.6Å(bp)^{-1}$, a 3% decrease). Although the imaging at zero forces suggests that the helix of DAP molecules increases its pitch, this change is not preserved when a low-to-moderate tension is applied. Considering also the CD spec-

troscopy results, one explanation for the discrepancy is that an unperturbed DAP molecule adopts a non-canonical conformation, but quickly reverts back to B-DNA when mildly stretched.

If DAP-DNA more readily adopts non-canonical forms at zero-tension it would be reasonable to expect those steric clashes to also facilitate conformational changes at high-tension, i.e. overstretching transition. For WT-DNA the characteristic overstretching transition around 60 pN is caused by different factors including force-induced melting and adoption of the S-DNA conformation [251]. The force-induced melting tends to originate at AT-rich regions while GC-sequences tend to form S-DNA. The melting and hysteresis experiments shown in the previous section demonstrate that DAP-DNA base pairing is significantly more stable than WT-DNA, presenting a significantly lower transition tension than WT sequences. This indicates that DAP-substitution not only disrupts B-DNA at low-tension but also lowers the energetic barrier between B- and S-form DNA.

11.1 CONCLUSION

Diaminopurine is a nucleobase analog of adenosine causing a third hydrogen bond when base paired with thymine. This additional amine resides on the minor-groove side of the DAP-T base pair and it not only stabilizes DAP-T pairs but significantly changes dsDNA conformation and mechanics. In this part, through AFM imaging, mechanical characterization using a MT, and CD spectroscopy it has been showed that DAP-substitution induces DNA molecules to adopt a non-canonical helix when in relaxed configuration and resist flexural bending at low-to-moderate tension (<10 pN), yet it also decreases the overstretching threshold from roughly 60 pN to 52 pN. From an energetic point of view, the steric clashes of DAP's extra amino group penalize entropic bends (increasing the persistence length) but, also raise the conformational free energy of the B-form helix, facilitating axial extension and initiating the B-to-S transition at a lower tension.

Part IV

DIRECT OBSERVATION OF CONDENSIN MEDIATED DNA-COLLAPSING

12

REAL TIME DETECTION OF DNA:CONDENSIN MULTISTEPS COLLAPSE

In the last ten years the use of single molecule force spectroscopy for studying biologically relevant processes has led the community to face the problem of thermal noise embedded into the single measurement. Indeed, the energy of thermal fluctuation (k_BT) is in the same order of magnitude of most biological events and it could not be avoided (i.e. it is not possible to lower the temperature or considering instrumental strategies to decrease the intrinsically thermal energy). This problem could be afforded by acquiring a relevant number of single traces and mediate the final results, or by applying some kind of filter (both in real time during the acquisition and *a posteriori*) or by using low temperature devices for recording the signals (CCD cameras). However, to clearly characterise some fast and sudden processes, it is necessary to work with raw and high-frequency-sampling data and to distinguish between the real events and the thermal fluctuations [252, 253]. This is the case of DNA-fast-condensation, as the one induced by SMC proteins, in particularly condensin [254].

In this part of the Thesis I afford the problem by adapting a MATLAB-written code to detect steps, smaller than the thermal noise, embedded into a extension-vs-time Magnetic Tweezers trace. To confirm the results, several instrumental test had been performed to determine the resolution of the instrument at the different employed forces (Appendix E).

This work has been done in the Moreno-Herrero Lab, CNB-CSIC, Madrid under the supervision of Dott. Fernando Moreno-Herrero and collaborating with Dr. Silvia Hormeño-Torres, Dr. Julene Madriarga-Marcos and Dr. Clara Aicart-Ramos. The condensin protein has been purified in the Luis Aragón Lab, Imperial College, London (see Section 3.3).

13

MATERIALS AND METHODS: CONDENSIN-DNA INTERACTIONS

13.1 DNA SUBSTRATES FOR MT

The DNA substrates employed in this work are two different torsionally-constrained standard DNA: a 6.310 Kbp-long (2.14 μm) and λ/2 DNA, both of them were fabricated as reported in [43] by Dr. Clara Aicart-Ramos in Moreno-Herrero Lab.

Briefly, for both th e 2-μm long and λ/2 DNA molecules, CosR-tail and CosL-tail oligonucleotides were biotin tailed and the Xba-A oligonucleotide was digoxigenin tailed using Terminal Transferase (NEB) and either BIO-dUTP or DIG-dUTP (Roche). The modified oligonucleotides were purified using a Qiaquick nucleotide removal kit (Qiagen). N6-Mehtyladenine free λ DNA (NEB) was cleaved with XbaI, giving two 24 508 bp fragments. These fragments and the three tailed oligonucleotides in addition to the XbaI-B oligonucleotide were subsequently annealed and ligated overnight using T4 DNA ligase (NEB).

13.2 MT FLOW CELLS

In this work, we employed a standard flow cell (with one inlet and one outlet) as described in [43, 44, 45]. Briefly, two silicate glass coverslips (Menzel-Gläser, # 1) were cleaned by 30 min of sonication in acetone followed by 30 min in isopropanol, and dried using compressed air. The clean surface of the lower part of the flow cell was coated with 1% polystyrene dissolved in toluene. The top cover glass contained two holes drilled with a laser engraver (VLS2.30, Universal Laser Systems), while the paraffin wax (Parafilm M, Bernis USA) gaskets were manually cut, to obtain a flow cell of ~ 100μl in volume. The two cover glass coverslip and the parafilm gasket were sandwiched and heated up for a few seconds at 120° to assemble the final flow cell. The cells were then incubated with an Antidigoxigenin (25ng/μl) solution (Roche) overnight at 4° and stored in a humid and sealed container at 4° until further use.

13.3 CONDENSIN PURIFICATION

The condensin protein has been purified in the Luis Aragón Lab, Imperial College, London following the procedure described in [54, 55, 58]. The samples were aliquoted and stored at $-80°$ until further use.

13.4 MAGNETIC TWEEZING

In this work, a home-made magnetic tweezers setup described in [46] was used. In brief, images of 1 µm streptavidin-coated superparamagnetic beads (Invitrogen, Carlsbad, CA) tethered to the surface of a glass slide by DNA constructs are acquired with a 100x oil immersion objective and a CCD camera. Real-time image analysis was used to determine the spatial coordinates of beads with nm accuracy in x, y and z. A step-by-step motor located above the sample moves a pair of magnets allowing the application of stretching forces to the bead-DNA system. Data were acquired at 120 Hz.

To obtain a homogeneous population of DNA molecules in the three configurations (Nicked, Torsionally constrained and Double), the DNA substrates were incubated 1 µm streptavidin-coated beads for 9 minutes with a gentle agitation every 3 minutes, then 10µl of free Biotin (Sigma) 1mM for 5 minutes for passivizate the beads without bounded DNA. Before flowing the sample into the flow chamber, the measurement buffer (50 mM TRIS−HCl pH 7.5, 100 mM KCl, 2.5 mM $MgCl_2$, 1 mM DTT, 0.1 mg/ml BSA) was added to the DNA:beads:biotin mixture to reach the total volume of 100µl.

Then, the sample was injected in the liquid flow cell the and incubated for 10-15 min before applying force. In a first preliminary step, visual inspection allows identification and selection of tethered DNA molecules. Torsionally-constrained molecules (TC), beads with more than a single DNA molecule (Double) and non torsionally-constrained filaments (Nicked) were identified from their distinct rotation-extension curves (for details see Fig.14.1).

13.5 DATA ANALYSIS - CODE

To assessed if the data are characterized by a step-like behaviour, I used two different user-independent step-finding algorithm, both of them developed in MATLAB language. The raw data were processed with Origin program (to subtract the reference bead trace, in order to avoid thermal shifting, and to subtract the zero of each molecule, to have the real DNA-extension on y-axes) before being analysed by means of the MATLAB code.

13.5.1 *Step-Finding-Algorithm*

The first algorithm, the *Step-Finding-Algorithm-1*, was basically implemented in MATLAB following the same work-flow described by *Kerssemakers et al* [252] and more recently used by Dekker and co-workers for analysing DNA:condensin MT traces [254].

The method works as follow: the data are cut (if necessary) to have the appropriate region of interest (i.e. - only the data at a certain force or after the flow stop). If there are some evident spikes in the trace they could be easily removed by a very short cycle to find and discard the maximum(s). Depending on the analysis, a gaussian-filter could be added to the data (where not otherwise specified the *smooth* was intended on 10-points gaussian-smooth filter).

The step finding was performed taking advantage of both the MATLAB function *findchangepts* and a custom written function *fitchangepts* in a typical line of code:

```
step_signal=fitchangepts
(data_10,findchangepts(data,'Statistic','mean',...
   'MaxNumChanges',n_step_iter1),'std');
```

The number of steps were routinely changed from $N_{stepiter} = 1$ to, namely $N_{stepiter} = 200$.

For each iteration, the χ^2 was independently calculated according to the equation 13.5.1:

$$\chi^2(N_{step-iter}) = \sum_{i=1}^{N_{sample}} (x_i - x_{step,i})^2 \tag{13.1}$$

Where N_{sample} was the number of points in the sample, x_i i-th points in the sample data, $x_{step,i}$ i-th points in the found step signal. The χ^2 for a signal with a thermal gaussian noise σ and N_{sample} points should be in the order of $\chi^2_{GoodFit} \sim N_{sample} \cdot \sigma^2$.

The step-like signal found was then slightly shifted (shift = 250 points, $\sim$ 2 s, signal-counter) and the $\chi^2_{counter}$ was calculated again between the data trace and the signal-counter. In theory, for a non-step like signal, a voluntarily mis-placed step-fitting should not affect the χ^2 result.

Therefore, the ratio between the χ^2 value of fit and counter-fit $\chi^2_{counter}$ reached a maximum at the most likely number of steps, whereas this ratio is 1 if there were no steps. The S-value was there plotted against the number of step finding. It shows a clear two-slope increase, and saturated for $N > N_{optimalfit}$. To highlight the transition step the first derivative (MATLAB function *polydif*) was calculated and the maximum was chosen for computing the best fit.

For particularly complicated traces this processed could be automatically repeated at least twice, deliberately choosing $N_{step} <$

$N_{optimalfit}$ and subtracting the resulting step-fit from the raw data. The second cycle was then run on the signal difference and the final $step - signal$ should be the sum between the $step - signal(1^{st}iteration)$ and the $step - signal(2^{nd}iteration)$.

13.5.2 *Step-Finding-Algorithm-optimized*

The second Step-Finding-Algorithm was provided by prof. Salapaka [253] and slightly adapted in MATLAB. Basically, this algorithm was an implementation of the one developed previously by Kerssemakers et al. [252], adding some feature here explained.

$J(x)$ *parameter*

The goodness of the fit was not evaluated by means of a ratio between χ^2 as before, but a new parameter J was defined as

$$J(x_i) = \sum_{i=1}^{N_{sample}} (x_i - x_{step,i})^2 + W \cdot \sum_{i=1}^{N_{sample}} \delta(x_i - x_{step,i}) \qquad (13.2)$$

where $\chi^2 = \sum_{i=1}^{N_{sample}} (x_i - x_{step,i})^2$ was the χ^2 error, $\delta(x_i - x_{step,i})$ was the Dirac delta function and $W \cdot \sum_{i=1}^{N_{sample}} \delta(x_i - x_{step,i})$ was the penalty introduced for having a new step. It was equal to have penalty $W \cdot N_{step}$ where $W = 9\sigma^2$ and N_{step} was the number of fitted steps. The minimum of J correlated with the best step fitting.

Probe Dynamics

The dynamic of the probe was included into the algorithm when the candidate $\hat{x}|J(\hat{x}) = min(J(x))$ calculating by a standard optimisation technique.

Evaluating the GoF - the LLR parameter

For evaluating the Goodness of the Function *GoF* and distinguishing between step-like and non-step-like signal, a new parameter had to be taken into account, the Log-Likelihood-Ratio (LLR) defined as:

$$LLR = \frac{1}{N}[\log P(x) - \log P(\tilde{x})] \qquad (13.3)$$

where $P(\tilde{x})$ was the probability of step fit x and $P(x)$ was the probability of smoothed fit $\tilde{x}$. According to [253] a LLR values > 0.08 indicated a stepping behaviour while an LLR < 0.05 indicated a non reliable fit.

14

RESULTS: CONDENSIN DRIVEN MULTISTEP DNA COMPACTION AND STEPS-FINDING SOFTWARE

14.1 CONDENSIN-MEDIATED DNA COLLAPSE

To measure the real-time compaction of DNA molecules by the *S. cerevisiae* condensin holocomplex we used a standard MT setup. We routinely performed a pre-measurement to determine the end-to-end extension of the bare DNA. The first step consisted in determining the *zero* of each DNA molecule by recording DNA fluctuations in absence of force (data not shown). Then, the protocol for distinguish between nicked (N), torsionally constrained (TC) and double DNA molecules was applied, as extensively explained in Fig. 14.1. Briefly, single nicked DNA molecules show no change in extension with rotations, but they display a decrease in the end-to-end distance due to the lower force. Single torsionally constrained DNA molecules do not display a length decrease at negative turns at high force but do form plectonemes at low force. Their mechanical response is different from that of double tethers, whose extension decreased forming plectonemes both with positive and negative turns, since the DNA molecules are entangled.

Then, condensin (30 - 40 nM) and ATP (1mM) were introduced simultaneously in the sample by flowing at a very slow rate (Q~20 µl/min). Following a short lag time, which probably reflects the time employed by the condensin to reach the DNA molecule and/or for interacting, the end-to-end length of the DNA started to decrease. In the vast majority of cases, as a consequence of the condensin interactions, we observed that the bead position moved closer to the surface (Fig.14.2, first line). We thus observe condensin-driven DNA compaction in real-time at the single-molecule level, in a very similar protocol as the one described in [254]. Decompaction of the DNA flow molecule could be observed by increasing the applied force of by adding high-concentration salt mixture (Fig.14.2, second line).

In this work, three different series of experiments were performed:

- 2µm-long DNA, F= 1 pN, Condensin 40 nM, ATP 1 mM
- 2µm-long DNA, F= 0.5 pN, Condensin 40 nM, ATP 1 mM

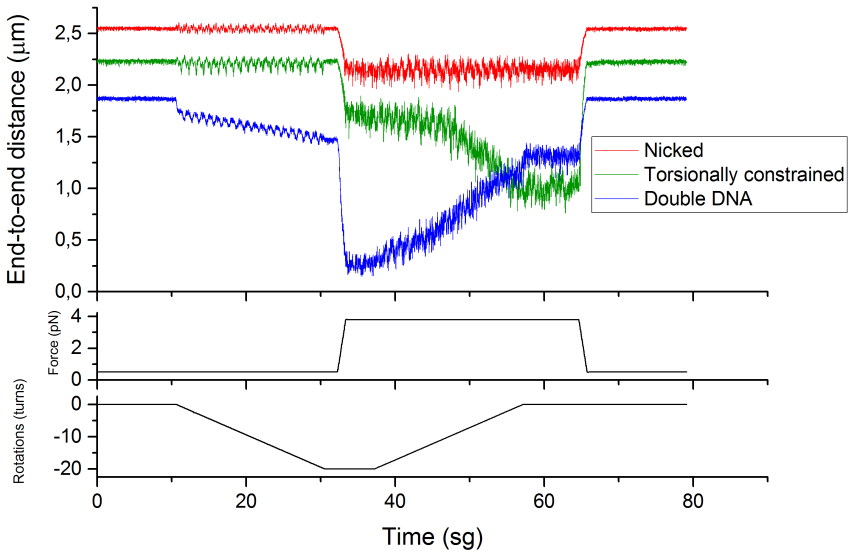

Figure 14.1.: Single nicked (red), single torsionally-constrained (green) and double (or multiple) DNA tethers (blue) were identified before each experiment by performing extension versus magnet turns curves at high and low forces.

- 8.6µm-long DNA, F= 0.75 pN, Condensin 30 nM, ATP 1 mM (reported in Appendix F)

A complete compaction of DNA molecules was usually obtained at forces between 2 and 0.5 pN (Fig.14.3). Double and TC molecules showed a faster end-to-end length decrease with respect to the nicked DNAs. Data at different forces showed a slower compaction at higher forces (Fig.14.4) and a minor thermal noise. Many DNA:condensin compaction traces showed sudden distinct decreases in the DNA end-to-end length, to which we will refer as steps. Some steps were found also is the opposite direction, showing a partial height recovery.

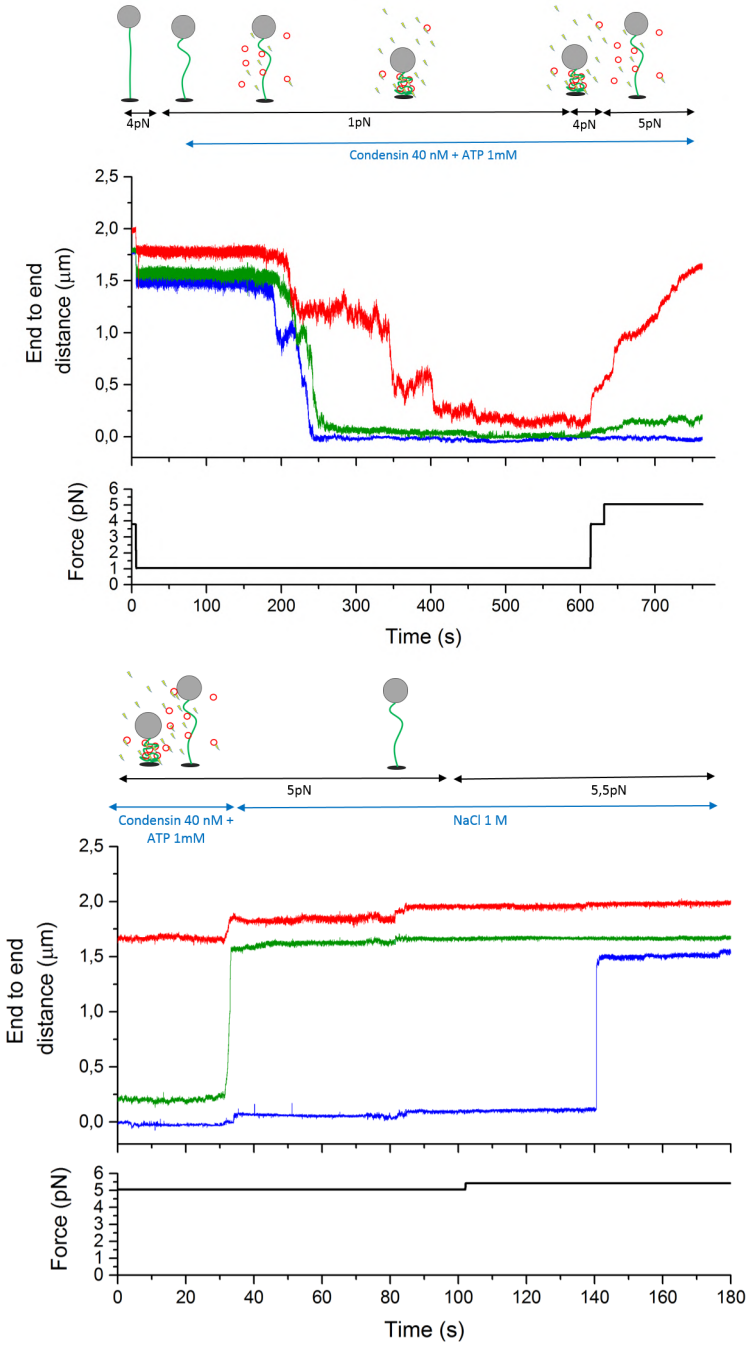

Figure 14.2.: (Line 1) The average DNA end-to-length was recorded at 4pN and 1pN. With the force at 1pN, condensin (30nM) and ATP (1mM) were added and the DNA was compacted. After some time, the force was suddenly increased to 4pN again. DNA end-to-end length increased, reversing compaction and eventually recovering the full end-to-end length.
(Line 2) The flow cell was washed with high salt (NaCl 1M), and the compacted structure was extended again. These traces were an example taken from several independent experiments for double (blue), N (red) and TC (green) DNAs. Although different tethers showed different rates of compaction and decompaction, the qualitative behaviours were identical.

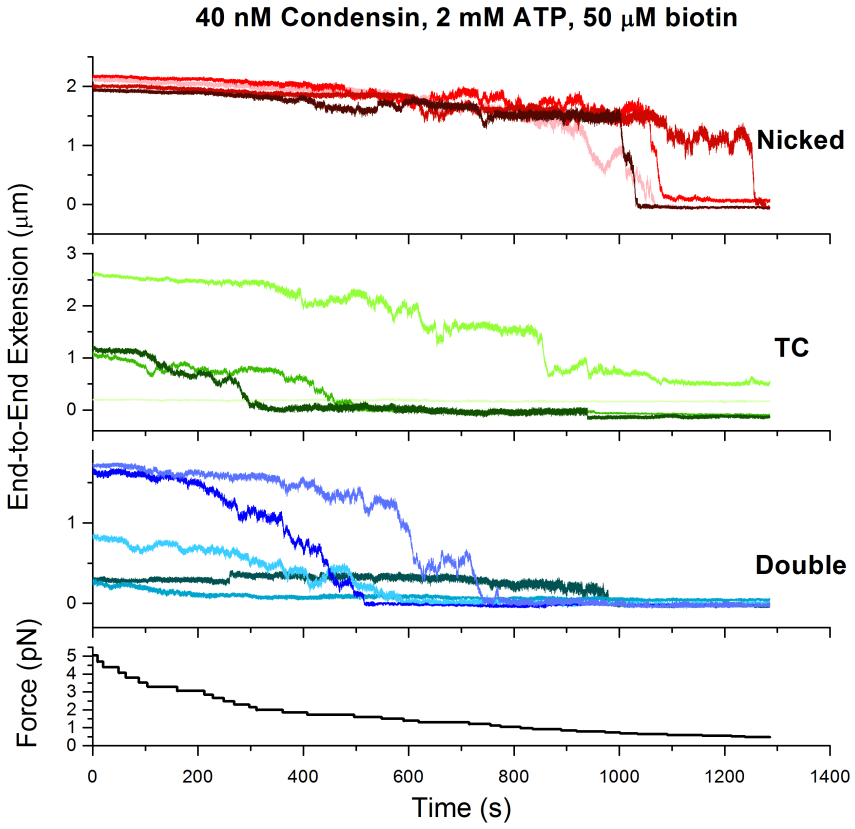

Figure 14.3.: DNA compaction in the presence of condensin 40 nM obtained by continuously decreasing the applied force. Nicked DNA molecules (first line, red) were totally compacted only at forces F < 0.5 pN, while TC and double showed a faster compaction rate.

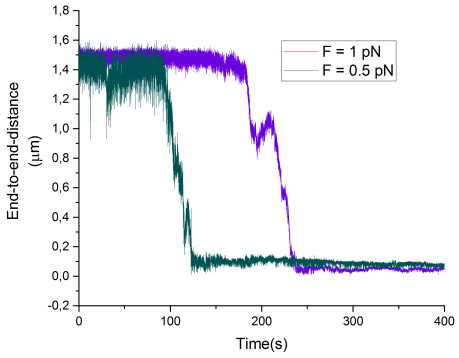

Figure 14.4.: Two example of traces of nicked DNA in the presence of 40 nM condensin, 1 mM ATP obtained at F=1pN (purple) and F=0.5 pN (dark blue).

The analysis of the steps in the DNA extension was particularly critical, due to the large thermal noise. Prior to apply the two distinct codes SFA and SFAO to our DNA:condensin data, several controls experiments had been done (see Appendix E for details).

14.2.1 *SFA validation*

Briefly, steps of different length were added to a real 2μm DNA trace by moving the objective. No differences were found at focus positions closed (±1μm) to the one usually employed in MT experiments. No differences were noticed also in alternated protocol (one step in one direction was always followed by another one in the opposite direction) or random up and down steps. Due to the intrinsic length of the condensin complex (~ 50 nm from the hinge to the heads) we focused on steps ranging from 100 to 30 nm. Briefly, at the forces considered for in the DNA:condensin compaction the SFA algorithm found steps whose length was compatible to the one expected within experimental error at 1 pN (Tab. 14.1), while for a lower force (0.5 pN) the 30-nm-long steps were largely overestimated (Tab. 14.2). All the results in the Tab. 14.1 and 14.2 were obtained by SFA (gaussian smooth 10, 2-cycles) by grouping all the three different focal positions.

Table 14.1.: **Force 1 pN** - SFA - Smooth 10, 2-iterations cycle

StepSize (nm)	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
100	-79.1 ± 54.2	77.9 ± 28.1	11	2072(188)
50	-46.7 ± 20.7	36.7 ± 19.8	11	2143(195)
30	-32.9 ± 13.3	22.4 ± 12.8	11	2089(190)

Table 14.2.: **Force 0.5 pN** - SFA - Smooth 10, 2-iterations cycle

StepSize (nm)	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
100	-72.2 ± 41.0	62.7 ± 39.0	7	1340(191)
50	-61.9 ± 28.9	50.6 ± 27.6	9	1474(164)
30	-55.62 ± 29.8	44.0 ± 27.4	9	1154(128)

14.2.2 *SFA on condensin:DNA traces*

In Fig.14.5 three typical examples of DNA compaction at 1 pN with fitted steps (coloured flat lines are the final steps, dot lines represent the steps found during the first iteration). We used this hands-off algorithm to analyse all traces we had collected to determine the size of the compaction steps.

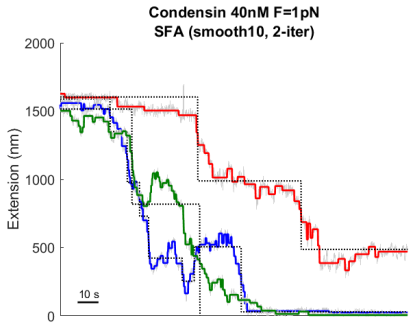

Figure 14.5.: Representative curves of condensin-mediate DNA collapse (F=1pN, 40nM, 2mM ATP) an step-find calculated by means of SFA (smooth 10, 2-iteration cycle). Double (Blue), Nicked (Red) and Torsionally Constrained (Green) DNA molecules. Dotted Black lines are the steps found by the 1-iteration Code.

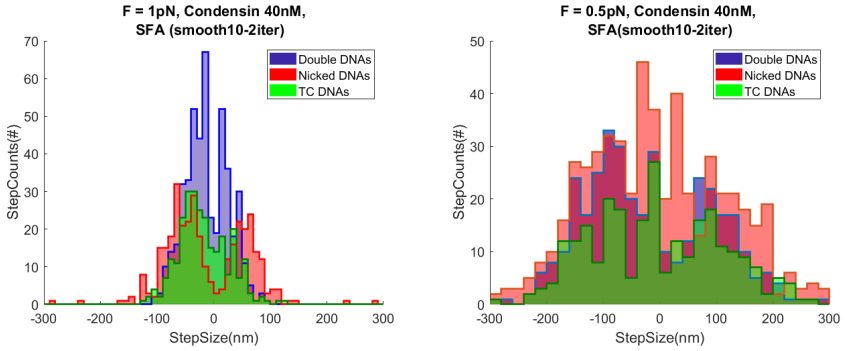

Figure 14.6.: Step-Size distributions of condensin-mediate DNA collapse (40nM, 2mM ATP) calculated by means of *SFA Code* (2-iteration cycle). Top: F=1pN, binSize=10nm, Bottom: F=0.5pN, binSize=20nm.

Fig.14.6 shows histograms of the distribution for the step size obtained at F=1 pN (top) and F=0.5 pN (bottom) divided for the DNA different typologies (N, TC and D). These two histograms displayed two distinct peaks, one of them corresponding to compaction steps (negative step values) and the other describing the less abundant spontaneous de-compaction steps (positive step values).

These results were then fitted with a double gaussian function, founding two different values (one for compaction and one for de-compaction steps). The traces relative to D, N and TC molecules did not show any significant difference among them (Tab. 14.5).

Table 14.3.: **Force 1 pN**

(nm)	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
Double DNAs	-29.4 ± 48.14	32.9 ± 18.5	5	478(96)
Nicked DNAs	-65.02 ± 44.0	46.7 ± 33.2	4	347(87)
TC DNAs	-48.9 ± 35.1	15.9 ± 35.0	3	279(93)
All DNAs	-44.5 ± 46.7	33.9 ± 33.5	12	1091(91)

Table 14.4.: **Force 0.5 pN**

(nm)	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
Double DNAs	-93.2 ± 95.0	90.1 ± 61.5	4	376(94)
Nicked DNAs	-40.4 ± 113	125.9 ± 44.0	4	370(93)
TC DNAs	-160.8 ± 54.4	124.8 ± 49.3	1	87(87)
All DNAs	-79.6 ± 112.6	97.0 ± 79.0	9	833(93)

Table 14.5.: Final step size distributions for DNA condensation in the presence of 40 nM condensin, 1 mM ATP found by SFA (smooth 10, 2-iterations cycle) software.

14.2.3 *SFAO on condensin:DNA traces*

We performed the same analysis also by using the SFAO, obtaining step size closed to the SFA-results (Fig.14.7 and Tab.14.8). We chose to apply the SFAO software to 10-point gaussian smoothing, without an indication on the value of the *measnoise* parameter (indicated here as SFAO(smooth-10, freenoise)). In Fig.14.11 detailed analysis also for SFAO(freenoise) on raw data were reported.

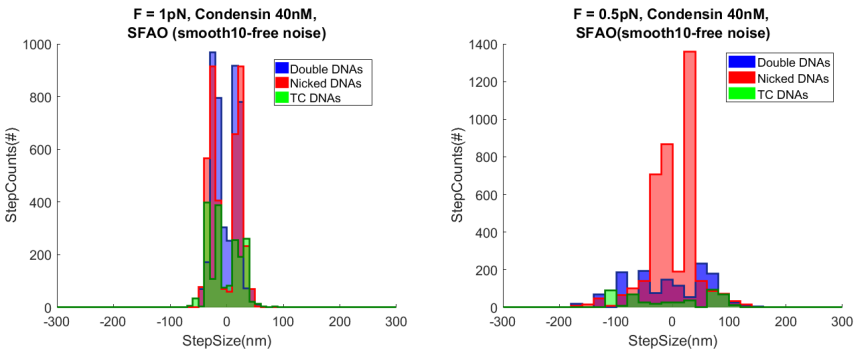

Figure 14.7.: Step-Size distributions of Condensin-mediate DNA collapse (40nM, 2mM ATP) calculated by means of *SFAO* (smooth-10, free noise). Top: F=1pN, binSize=10nm, Bottom: F=0.5pN, binSize=20nm

Table 14.6.: **Force 1 pN**

(nm)	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
Double DNAs	-35.3 ± 22.5	21.7 ± 17.6	5	1146(229)
Nicked DNAs	-40.2 ± 17.9	27.1 ± 13.8	4	2199(550)
TC DNAs	-39.4 ± 32.3	37.3 ± 7.2	3	683(228)
All DNAs	-39.2 ± 19.6	26.6 ± 17.1	12	4028(336)

Table 14.7.: **Force 0.5 pN**

(nm)	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
Double DNAs	-85.3 ± 82.2	74.4 ± 26.9	4	866(217)
Nicked DNAs	-43.6 ± 18.2	32.5 ± 22.5	4	2090(523)
TC DNAs	-104.8 ± 164.4	122.6 ± 52.2	1	224(224)
All DNAs	-46.7 ± 22.3	42.4 ± 54.5	9	3180(353)

Table 14.8.: Final step size distributions for DNA condensation in the presence of 40 nM condensin, 1 mM ATP found by SFAO (smooth-10, freenoise) software.

Table 14.9.: **Force 1 pN**

(nm)	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
Double DNAs	-46.9 ± 31.3	7.7 ± 30.7	5	1146(229)
Nicked DNAs	-48.8 ± 22.1	18.3 ± 25.8	4	2199(550)
TC DNAs	-56.2 ± 30.7	22.9 ± 29.5	3	683(228)
All DNAs	-49.6 ± 26.9	16.6 ± 27.9	12	4028(336)

Table 14.10.: **Force 0.5 pN**

(nm)	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
Double DNAs	-104.4 ± 76.4	59.9 ± 31.5	4	866(217)
Nicked DNAs	-58.9 ± 22.0	19.8 ± 38.0	4	2090(523)
TC DNAs	-111.6 ± 161.5	116.2 ± 72.2	1	224(224)
All DNAs	-59.0 ± 33.4	20.0 ± 73.2	9	3180(353)

Table 14.11.: Final step size distributions for DNA condensation in the presence of 40 nM condensin, 1 mM ATP found by SFAO (freenoise) software.

14.2.4 *SFA vs SFAO on condensin:DNA traces*

The number of steps detected by the SFAO software were sensibly higher than the ones revealed by the SFA algorithm (~ 340 for each trace for SFAO and ~ 92 for SFA). Despite this high discrepancy in the number of found steps, for cleaner data at higher force (F = 1pN), the size of detected steps were compatible within the error in both the compaction (negative values) and de-compaction (positive values) directions.

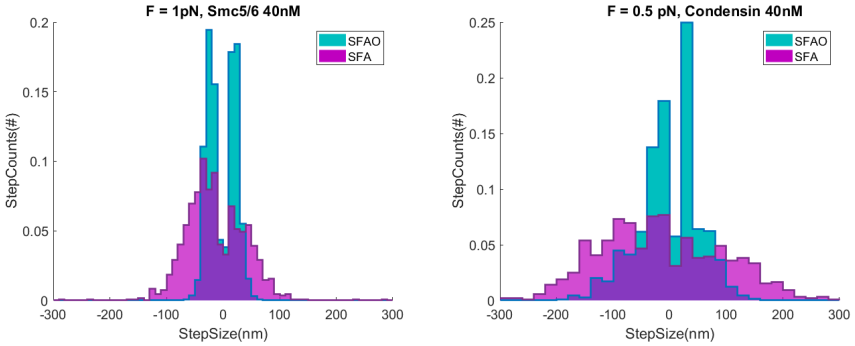

Figure 14.8.: Step-Size distributions of condensin-mediate DNA collapse (40nM, 2mM ATP) calculated by means of *SFAO* (smooth 10, free noise, Light Blue) and *SFA* (smooth 10, 2-iter cycle , Purple). F=1pN, binSize=10nm, F=0.5pN, binSize=20nm.

Unfortunately, the same results were not found in data obtained at 0.5 pN, probably due to the high thermal noise.

Table 14.12.: **Force 1 pN**

(nm)	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
SFA	-44.5 ± 46.7	33.9 ± 33.5	12	1091(91)
SFAO	-39.2 ± 19.6	26.6 ± 17.1	12	4028(336)

Table 14.13.: **Force 0.5 pN**

(nm)	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
SFA	-79.6 ± 112.6	97.0 ± 79.0	9	833(93)
SFAO	-46.7 ± 22.3	42.4 ± 54.5	9	3180(353)

Table 14.14.: Final step size distributions for DNA condensation in the presence of 40 nM condensin, 1 mM ATP found by SFA (smooth-10. 2-iteration-cycle) and SFAO (smooth-10, freenoise) software.

15

DISCUSSION: CONDENSIN MULTISTEP COMPACTION OF DNA IN A TENSION-DEPENDENT MECHANISM

15.1 STEPS FINDING ALGORITHMS

In this study, two different steps-finding algorithms were partially implemented and validated on real DNA traces. We decided to not introduce simulated false steps in the traces in order to be as close as possible to real experimental conditions, thus we recreated a stairs-like signal by physically moving the piezoelectric.

This process was more subject to environmental noise and to problems during the tracking, both on the DNA beads and on the reference signal. It was also affected by changes in the focal position and mechanical-related shifts. For all these reasons, we were quite confident to not overestimate the goodness of the software and to validate it with traces which reflected the real conditions of the condensing DNA molecule. Two different sources of noise in the extension-vs-time trace could be individuated. Firsts of all, to introduce proteins and ATP, it was necessary to work in an *open* flow cell: the outlet was connected to a syringe pump while the inlet was set free to permit the injection of new and/or different solution. This configuration was clearly less stable than a completely insulated system. Moreover, the noise in the extension traces was strongly affected by both the applied magnetic force and the length of the DNA molecule (i.e, the longer was the thread pendulum the more it could fluctuate). While the force was usually kept constant (at values between 0.3 and 1 pN) both on SFAs validation tests and on the real DNA traces in the presence of the condensin, the different extension of the DNA was difficult to be taken into account.

Indeed, during the condensin mediated DNA compaction, the molecule was reducing its length, and, simultaneously limiting its free oscillations, which were translate in a lower noise. This problem could affect also the identification of the condensation steps by the algorithm. Actually, it enhanced steps in the first part of the trace, when the standard deviation of the signal is higher. Unfortunately, all the solutions tested, i.e. segmentation of the trace or a different smoothing of the raw data depending on the extension, were not applicable

to fast condensation data without affecting also the steps embedded into the trace. For this reason, we decided to focus on the entire trace and apply the algorithms on both raw and smoothed data.

As shown in Appendix E, we used 2µm-long DNA molecules to recreate the exact conditions of the real measurements for all the test performed. At the maximum force reached by the setup (~ 5 pN) the SFA code was able to correctly resolve steps of 100, 50 and 30 nm. (Tab.E.53 for smoothed-by-10 data and E.52 for the raw signal). Unfortunately, the condensin was not able to act on DNA in the presence of a such high force.

For the forces considered in the Chapter 14 for the condensin data, the SFA software applied to smooth-by-10 data was able to resolve steps of 100, 50 and 30 nm (slightly underestimating the 100 nm steps) at 1 pN, while for the 0.5 pN traces the steps under 50 nm were baldly overestimated.

We applied also the SFAO routine to the same data, but the final number of steps found was severely (at least two times) higher than expected. It was also found that the positive values (corresponding to de-compaction events) were systematically smaller than their relative negative values.

For all these reasons, we decided to apply the SFA algorithm on smoothed DNA:condensin data at 0.5 and 1 pN forces. We decided also to not include all the data, previously obtained by Dr. Hormeño-Torres, for the $\lambda/2$ DNA because we did not perform the validation process in a such a long DNA molecule. Data of $\lambda/2$ DNA at 0.75 pN are reported in Appendix F.

15.2 INSIGHTS IN CONDENSIN-DRIVEN DNA COMPACTION

Unravelling the mechanism of SMC which drives the DNA into a compact and structurally defined final form is still a challenge. In the last few year a lot of different models were proposed, most of them speculated from the results obtained by single molecule characterisation. However, there is not yet an unique model which can explain all the features presented by the SMC proteins, as well as there is not a complete agreement regarding the differences of the SMC complexes among them. Considering the three eukaryotic SMC complexes, condensin, cohesin and SMC5/6, it is known that they fulfill different functions, which may also partially overlap. Recently, the model proposed for the cohesin protein [202] is not totally super imposable to the ones proposed for the condensin protein [54, 55, 58].

One of the most important requirements for clarify the compaction mechanisms driven by the SMC complexes is to establish if they progress by fixed steps, by steps of different extensions or they condense in a continuous process. As specified above, this is particularly difficult to determine for intrinsically noisy single molecule data.

In [254] it was pointed out that the SFA algorithm was not able to resolve steps smaller than 100 nm added to a simulated trace. We demonstrated that, for an applied magnetic force of 1 pN, our routine was able to resolve steps up to 30 nm occurring in both directions.

According to our validation processes of SFA, we can claim that the steps performed by condensin on DNA at 1 pN were $\sim$ 50 nm while for lower force (0.5 pN) the results were compatible with $\sim$ 100 nm steps. Obviously, we can not completely eliminate the possibility of smaller steps which were not detected, but we can rule out that, due to different step sizes found at different forces, the condensin protein does not act for fixed steps.

Among all the models proposed (and discussed in Section 3.3.5) we can eliminate the *inchworm model* [201], which required a fixed steps of $\sim$ 50nm for all the forces considered and the *sequential walking model*, which requires sensibly smaller steps [200]. All the *loop extrusion* model [54], the *random cross-linking* model [200] and *segment captured model* [201] are compatible with our results, as they do not need steps of fixed height for the condensation process.

We did not find any differences between the three DNA configurations studied (nicked, torsionally constrained and double DNA molecules) neither in the standard MT experiments (Section 14) nor in the presence of negative or positive rotations (data not shown).

We also found that for DNA molecules of different size (2µm, Chapter14 and 8.6µm, Appendix F) results are not totally super imposable. Indeed, we found higher steps for the long $\lambda/2$ DNA with respect to the shorter DNA at comparable forces. This probably is due to a condensation too fast to be totally resolved by both the tracking procedure and the SFA routine.

15.3 CONCLUSIONS

The results reported here visualize the interactions of condensin complexes with DNA. We were able to observe the condensation at different forces and with different DNA configurations. Our results reveal that the condensin compacts the DNA in a multistep mechanism, with a height of each step depending on the tension (force) exerted on the molecule. Importantly, our findings of condensation depending to the force are inconsistent with previous models that suggest condensin condense DNA in constant steps independent of the tension, like the proposed *inchworm model* and instead agree with models in which condensin extrude portions of DNA consistently larger than the complex itself.

Part V

DNA FLOW STRETCHING COUPLED WITH SINGLE MOLECULE FLUORESCENCE

16

FLOW STRETCHING INSTRUMENTAL DEVELOPMENT FOR DNA TETHERING

Single-molecule studies usually take place in flow cells. This approach is not only necessary for experiments which require a liquid environment, but is also useful to allow the exchange of reagents before or during measurements. This is crucial in experiments that need to be triggered by ligands, ATP, or reactions that require a sequential addition of proteins. The achievement of a very rapid buffer exchange is particularly useful in DNA:protein binding-unbinding experiments, as the ones reported in [43]. During my stay at the Moreno-Herrero Lab I was involved in a project consisting in the optimisation and validation of new microfluidics chambers for a TIRF-based flow stretching apparatus for the visualisation of stretched DNA and its interactions with different proteins. All the measurements reported here have been performed on the MT-TIRF setup described in [43, 44, 62] and realized by Dr. Moreno-Herrero and Dr. Madriarga-Marcos.

I employed simple home-fabricated multistream laminar microfluidic cells (from here, flow cells) consisting of two glass coverslips and a gasket made of paraffin wax (parafilm) in the middle, analogous to the flow cells used for standard MT experiments (Section13.2). The volume and the shape of each flow cell could be easily controlled by modifying the two dimensions of the channel (both height and width).

In a typical Flow Stretch (F-S) experiment, the reagents were introduced using a system of syringe pumps. In such a system, a fast and controlled exchange of reagents was not possible without working under very high flow rates, which disturb both the biological sample and the instrumental setup itself, impeding performing sophisticated experiments due to the intrinsic instability of the system. On the other hands, lower flow rates, employed to not perturb the system, usually lead to the generation of a reagent gradient in the flow cell. For very precise measurements it is thus desirable to have a very fast exchange between solutions with minimal diffusion.

The MT-TIRF flow stretching apparatus can be essentially employed in two different modalities: (1) to apply a stretched force to a DNA molecule anchored to a surface and imaged via a fluorescence

illumination or (2) to achieve a rapid exchange of buffer, i.e. changing ligands, buffer or proteins, during a standard MT experiment. These two requirements, which are partially in conflict, can be obtained by improving the design of the flow cells, in order to permit a fast buffer exchange even in the presence of slow flow rate.

We proposed an implementation of the standard flow cells (consisting of two inlets and one outlet) reducing the width of the central channel from 7 to 1 mm. In this scenario, because of the nature of laminar flow, a fast exchange was achieved while diffusion was minimized. Reagents were introduced by two computer-controlled syringe pumps. Our two-inlet flow cells were home-designed and fabricated, and their performance had been characterized by MT and TIRF experiments.

In the first characterisation experiment, the buffer exchange in different channels designs was evaluated by recording the fluorescence near the surface. We took advantage of TIRF illumination to characterize the boundary switching between 1 mM Fluorescein and buffer in several flow cells with different cross-sections (Section 18.1).

In the second chapter, a simple fluorescent DNA-stretching experiment was performed in the narrower flow cell: Sitox green labelled DNA was visualised under TIRF illumination at different flow rates and the DNA:Sitox binding was evaluated (Section 18.2).

Finally, in the last part (Section 18.3), we calibrated the force applied to a DNA molecule by the external flow, using a force-calibration protocol very similar to the one presented in [44]. Briefly, DNA molecules, with 1 μm standard bead attached to one end like a traditional MT sample, were extended by the external flow.

By taking advantage of the MT-related-CCD camera, we recorded the position of the DNA-coupled beads under different flow rates, correlating both the DNA extension and the force with the applied flow rate.

17

MATERIALS AND METHODS: FLOW STRETCHING

17.1 COMBINED FLOW STRETCHING-TIRF APPARATUS

Experiments were performed in a F-S - MT-TIRF setup similar to one described in [43, 44]. In brief, 488 nm laser source (Vortran Stradus) was focused on the back focal plane of a high numerical aperture objective (Olympus UAPON TIRF 100x). Two separate detectors were employed to visualise the emission of the fluorophores, an EM-CCD temperature-controlled camera (Andor Ixon Ultra 897/Andor Ixon Ultra 888) and a CCD or CMOS camera (Pulnix 6710 CL/Mikrotron MC1362) for bright-field video microscopy. This latter camera was controlled by a custom-written code [255] while fluorescence camera was controlled by Andor Solis software.

17.2 FLOW CELLS

In this work, the flow cells had a typical two-inlet and one-outlet configuration, with different channel width and height. In all the cases considered, glass coverslips (Menzel-Gläser, #1) were cleaned by 30 min of sonication in acetone followed by 30 min in isopropanol, and dried using compressed air. The top cover glass contained three holes drilled with a laser engraver, as well as two-inlet paraffin wax (Parafilm M, Bernis USA) gaskets (VLS2.30, Universal Laser Systems). The two cover glass slides and a (single or double layer) gasket were sandwiched and heated up for a few seconds at 120°C to assemble the flow cell, refrigerated and stored at room temperature. For DNA thetering experiments, the clean surface was coated with 1% polystyrene dissolved in toluene, before the flow cell assembly. The DNA-tethered cells were then incubated with an Antidigoxigenin (25 ng/µl) solution (Roche) overnight at 4°C and were passivizated for at least 2 h using BSA (NEB). The cells were stored in a humid and sealed container at 4°C until further use. For fluorescein characterisation, three different parafilm gaskets were used (with the central channel 1, 1.5 or 2 mm width) and one or two parafilm layers were employed to achieve an height of $\sim$ 100µm (1 layer) or $\sim$ 200µm (2 layers). For both DNA-tethering measurements and force calibration

(Section18.2 and18.3), the narrower flow cells (1 layer - 1 mm width) were employed.

17.3 λ DNA

λ molecules were fabricated as described elsewhere [44] by Dr. Aicart-Ramos. Briefly, CosR-tail and CosL-tail oligonucleotides were biotin tailed using Terminal Transferase (NEB) and BIO-dUTP (Roche). The modified oligonucleotides were purified using a Qiaquick nucleotide removal kit (Qiagen). N6-Mehtyladenine-free λ DNA (NEB) was cleaved with XbaI, giving two ~24 kbp fragments.

17.4 FLUORESCEIN MULTICHANNEL LAMINAR-FLOW EXPERIMENTS FOR RAPID BUFFER EXCHANGE

To have a clear focus on the surface, a very diluited sample of paramagnetic beads 1μm were flowed inside the flow channel before the measurements in MilliQ water. The speciment was abundantly clean with water before to perform the TIRF fluorescent measurements. Fluorescein (Sigma Aldrich) was dissolved in Hepes 50 mM, pH 8 at the final concentration of 1 mM. To achieve the boundary exchange measurements the fluorescein reservoir was intermittently switched with the buffer alone by using two automatised syringe pumps. Syringes were controlled with the neMESYS UserInterface software. Briefly, a square-wave pattern was set for syringes to alternate flow-rates, keeping a constant flow-rate in the central channel (Q from 7 to 200 μl/min). Fluorescence images were acquired using Andor Solis software, at a frequency of 10 Hz for a tot 500 frame for each tested flow rate. All the measurement were carried out in the central channel of the flow cells, approximately at 1-2 mm from the two-inlets intersection. Data analysis was performed using Andor Solis software, ImageJ and Origin. For each flow rate, at least 5 regions of interest (ROI) of ~ 50x50 pixels were selected in an homogeneous area and their fluorescence emission analyzed in Origin software.

17.5 DNA-TETHERING MULTICHANNEL LAMINAR-FLOW EXPERIMENTS

Tethers of λ DNA were obtained by flowing DNA molecule in standard buffer (50 mM Tris pH 7.5, 100 mM KCl, 2.5 mM $MgCl_2$, 1 mM DTT and 0.1 mg/ml BSA). After biotin-streptavidin bound were formed, unbound DNA molecule were extensively washed away. To visualized DNA molecules under the TIRF illumination, the intercalant Sytox green was used. DNA molecules were stained with 100 nM Sytox Green in standard buffer supplemented with 1 mM Trolox, 20 mM glucose, 8 μg/ml glucose oxidase and 20 μg/ml catalase to

avoid the early photobleaching. Fluorescence images were acquired using Andor Solis software, at a frequency of 9.52 Hz, using the EM level of 100 and cooling the sensor to -60°C. At least 4000 frames were recorded. Syringes were controlled with the neMESYS UserInterface software. Briefly, a square-wave pattern was set for syringes to alternate flow-rates, keeping the flow-rate in the central channel constant (Q to 10-200 μl/min). Data analysis was performed using Andor Solis and Origin as described in [62]. For each DNA molecule of interest, a main region of interest (ROI) was selected around the DNA molecule (additional ROIs around the main ROI were also selected for background correction). Fluorescence intensities were then analyzed in Origin software.

17.6 FLOW STRETCH FORCE CALIBRATION

Tethers of λ/2 DNA molecules (see Sec. 13.1) were obtained by mixing DNA with 1 μm size magnetic beads (Dynabeads, MyOne streptavidin, Invitrogen) in a buffer containing 10 mM PB (pH 7), 10 mM NaN_3, 0.2 mg/ml BSA, and 0.1% Tween 20. Sample was then injected in an anti-digoxigenin functionalised 1 layer-1 mm channel-width flow cell and incubated for 10-15 min before applying the external flow. In the preliminary step each DNA molecule was individually calibrated using the standard MT force calibration protocol. Then, the magnets were removed and the flow was initialised by using the two in-parallel-working automatised syringe pumps filled with the same standard buffer. Briefly, a step-like profile with increasing flow rates (from 2 to 150 ml/min, 5 s for each step) was initialised in the neMESYS software. The 3D position of each bead was recorded and then converted into DNA extension and force.

18

RESULTS: FAST BUFFER EXCHANGE, DNA STRETCHING AND FORCE ESTIMATION

18.1 FLUORESCEIN EXPERIMENTS: BOUNDARY EXCHANGE

18.1.1 *Boundary exchange at different cross section channels*

In single molecule studies, it is crucial to be able to switch different biochemical reagents, keeping the flow velocity low enough so that the cells and molecules remain attached and, at the same time, achieving the buffer exchange as fast as possible [59]. The shape of the flow cell, i.e. its cross section, is one the main feature which is involved in the boundary switch and could be implemented in a straightforward way, following the same protocol currently in use in Moreno-Herrero lab.

We constructed 5 flow cells, whose channel presented different heights and widths (Fig.18.1), using a 2-inlets-1-outlet design.

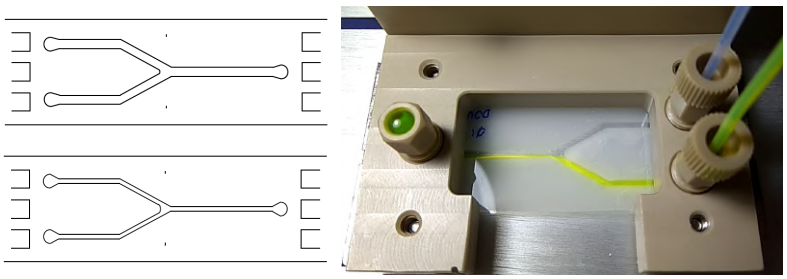

Figure 18.1.: (Left) Different designs for two channel microfluidic cells: 2 mm-wide (top) and 1 mm-wide (bottom) channels.
(Right) Picture of a 1 mm-wide flow cell mounted on the F-S stage for buffer:fluorescein experiments.

Taking advantage of simulations described in [59], we focused all the measurements in the proximity of the inlet (~ 1-2 mm) to achieve the fastest boundary switch. We used a square wave to alternating the injection of the two buffer with a switching time of $t_{switch} = 10$ and different flow rates, ranging from 7 $\mu l \cdot min^{-1}$ to 20 $\mu l \cdot min^{-1}$. Final flow rates were built as indicated in Tab.18.1.

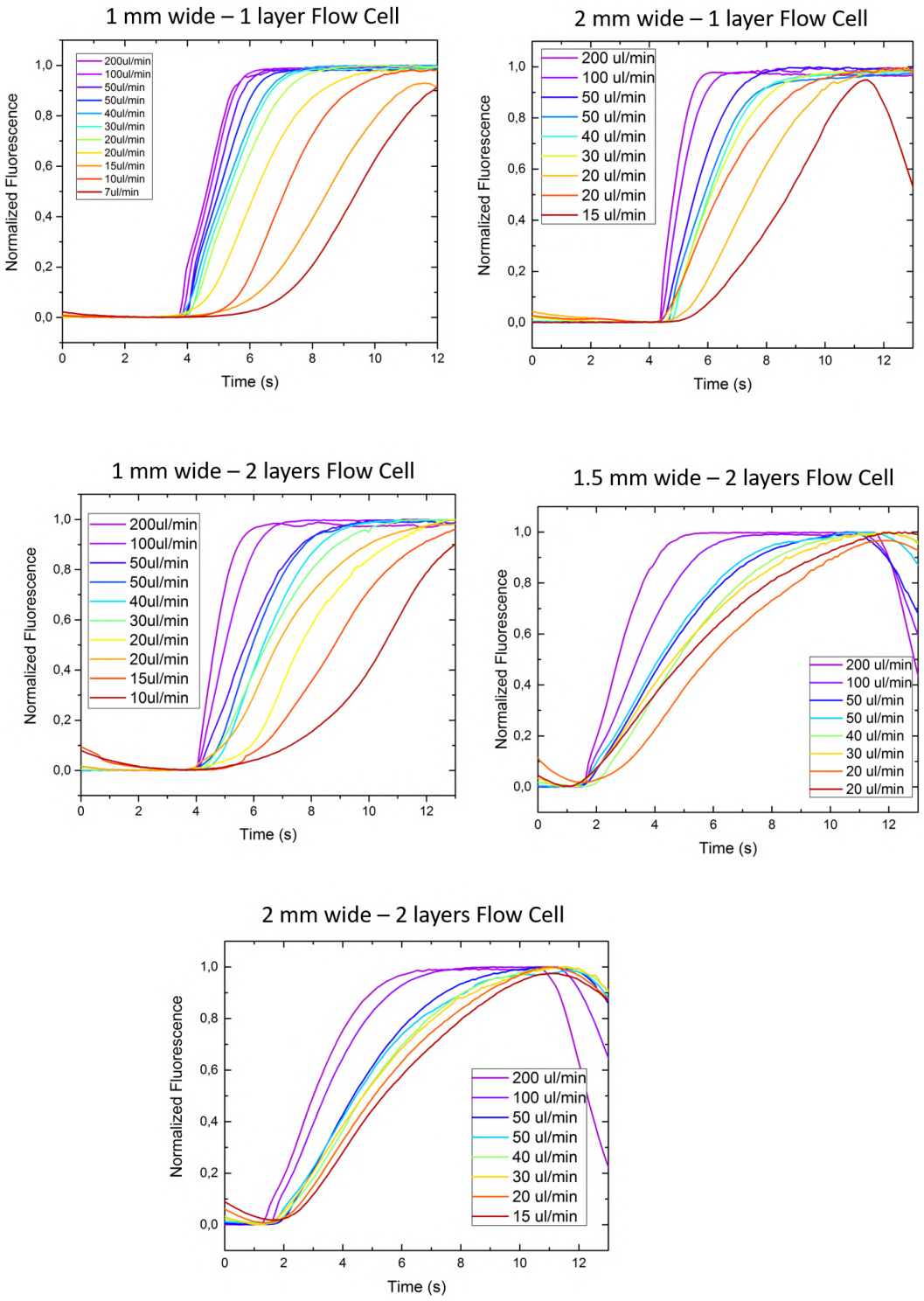

Figure 18.2.: Boundary Exchange of Fluorescein at different flow rates for five flow cells designs. From Top-Left: 1 mm wide - 1 layer, 2 mm wide - 1 layer, 1 mm wide - 2 layer, 1.5 mm wide - 2 layer and 2 mm wide - 2 layer flow cells.

Final Q ($\mu l \cdot min^{-1}$)	Maximum Q (Q_{max}) ($\mu l \cdot min^{-1}$)	Minimum Q (Q_{min}) ($\mu l \cdot min^{-1}$)
200	190	10
100	90	10
50_1	40	10
50_2	45	5
40	35	5
30	25	5
20_1	15	5
20_2	17	3
15	12	3
10	7	3
7	5	2

Table 18.1.: Final flow rates Q obtained as the sum of the two Q_{max} and Q_{min}. For the Q = $50\mu l \cdot min^{-1}$ and Q = $20\mu l \cdot min^{-1}$ two different compositions have been tested, usually showing very similar results. For flow cells with high cross section the buffer exchange was non achieved for the lower Q values.

To characterize the quality of switching between fluids [59], we defined the fall time t_{switch} as the time to go from 90% intensity to 10% intensity (Fig.18.3) [59, 44].

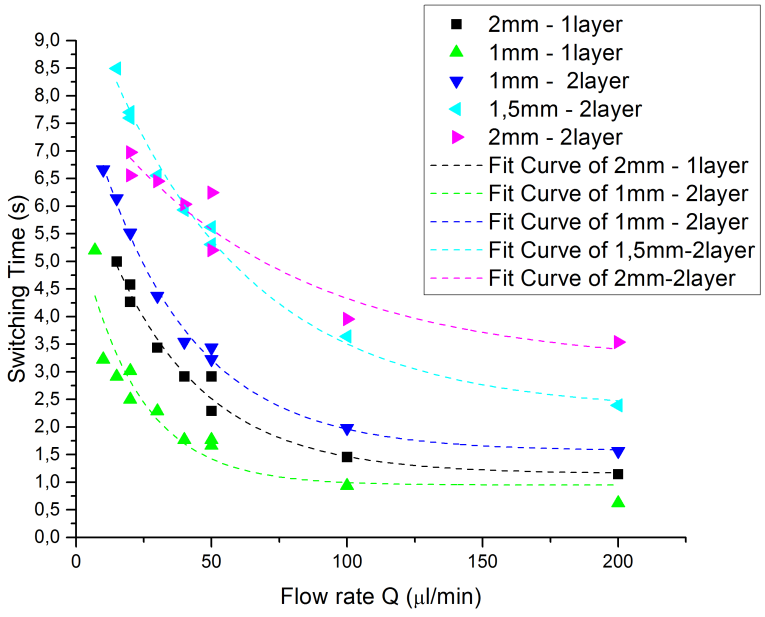

Figure 18.3.: Switch time t_{switch} calculated for different flow cells versus the total flow rate Q. Dashed lines represent the individual fit with an exponential decay function (Eq.18.1.1).

$$t_{Switch}(Q) = A \cdot e^{\frac{Q}{t_1}} + t_0 \tag{18.1}$$

The fitting parameter t_1 was linearly dependent to the channel cross section (Fig.18.4 and Tab.18.2)

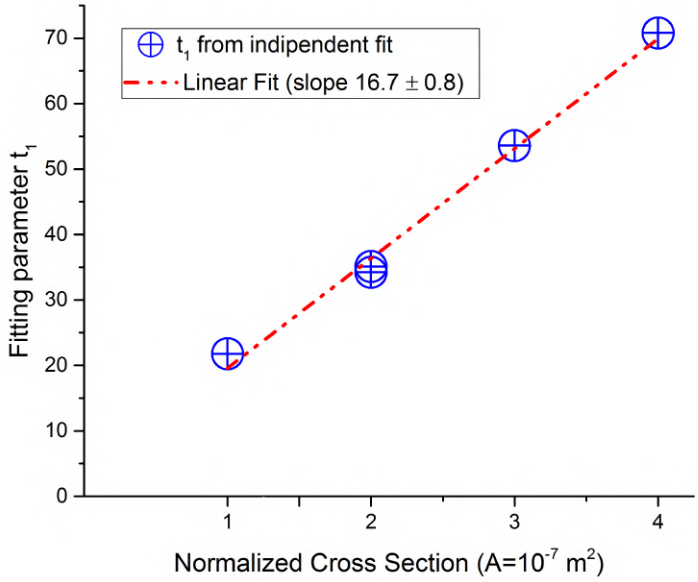

Figure 18.4.: The t_1 fit parameter as a function of normalized cross section A, where A is the cross section of the 1 mm - 1 layer flow cell channel ($A = 1 \cdot 10^{-3}[m] \cdot 1 \cdot 10^{-4}[m]) = 10^{-7} m^2$. Linear fit slope is 16.7 ± 0.8.

Flow Cell	Cross section (ϑ^2) (μm^2)	Fitted parameter (t_1) (min $\cdot \mu l^{-1}$)
1mm - 1 layer	A	21.76
2mm - 1 layer	2A	34.22
1mm - 2 layer	2A	35.08
1.5mm - 2 layer	3A	53.57
2mm - 2 layer	4A	70.78

Table 18.2.: The t_1 fit parameter as a function of normalized cross section A.

18.1.2 *Taylor-Aris model*

We fitted our fluorescence vs time signals to the Taylor-Aris model for a rectangular channel as described in [59, 60, 61] (for details see Sec.1.5.2):

$$c(t) = c_0 \frac{1}{2}\left(1 + \mathrm{Erf}\left[\frac{y - vt}{2\sqrt{tk}}\right]\right) \tag{18.2}$$

where c_0 is the concentration of fluorophores, v is the average linear velocity, k is the effective diffusion coefficient and y the distance from the junction. v and k should change with flow rates and between different cross sections while y should remain unchanged among the same flow cell.

Our flow cell designs fulfilled the requirements for the Taylor-Aris approximations (the following calculations were performed for our wider flow cell (2 mm wide - 2 layers)):

$$y_0 = \frac{a^2 v_0}{288D}$$

where y_0 is the minimum distance from the entrance of the interface, $v_0 = 160\mu m/s$ is the velocity of the fluid, and $a = 200\mu$ the height of the channel and $D = 4.90 \cdot 10^{-10} m^2/s$. The Taylor-Aris is valid at distances $y > y_0$ from the two-inlet junction (in our setup $y_0 = 0.004mm$ and $y =\sim 2mm$).

Results for Taylor-Aris fitted data are reported in Tab.18.3 as the average of independent fits (one signal for each flow rate Q).

Flow Cells	Cross section (μm^2)	y_0 (m)	D (m^2/s)	a (m)
1 mm - 1 layer	A	$8.71 \cdot 10^{-4}$	$2.08 \cdot 10^{-9}$	$5.81 \cdot 10^{-5}$
2 mm - 1 layer	2A	$2.01 \cdot 10^{-3}$	$1.46 \cdot 10^{-10}$	$1.33 \cdot 10^{-4}$
1 mm - 2 layer	2A	$6.30 \cdot 10^{-4}$	$1.54 \cdot 10^{-9}$	$1.28 \cdot 10^{-4}$
1.5 mm - 2 layer	3A	$1.85 \cdot 10^{-3}$	$5.13 \cdot 10^{-10}$	$1.02 \cdot 10^{-5}$
2 mm - 2 layer	4A	$1.2 \cdot 10^{-3}$	$3.18 \cdot 10^{-9}$	$4.28 \cdot 10^{-4}$
Theoretical		$1 \cdot 10^{-3}$	$4.9 \cdot 10^{-10}$	$1 - 2 \cdot 10^{-4}$

Table 18.3.: Independent Taylor-Aris fit for the five considered flow cells. 4 free parameters: y_0 (distance from inlet), v (fluid speed - not reported), a (channel height), D (diffusion constant). Data displayed here were the average of independent fits.

Global and independent fits of two representative flow cells are reported in Appendix G.

After having characterized different flow cell design, we employed the narrower one (1 mm wide - 1 layer) to characterize the behaviour of DNA under the influence of an external flow.

To visualize the DNA under TIRF illumination we took advantage of an intercalant, the Sytox green, which had been widely employed in DNA-fluorescence essays [52, 53, 54, 57].

The family of Sytox nucleic acid staining dyes had been commonly used as dead-cells markers, but recently, it is largely employed in single-molecule experiments. Sytox dyes exhibit a large (>1000-fold) fluorescence enhancement upon DNA binding [52]. Moreover, according to [52], the Sytox intercalant binds DNA rapidly and with high affinity, has a good signal-to-noise ratio even at low concentrations (in this work the Sytox was employed at 100 nM), exhibits a low photobleaching rate and induces lower light-induced DNA degradation.

Different flow rates were applyed to our sample, from 200 to 10 $\mu l \cdot min^{-1}$ (Fig. 18.5, 18.6, 18.7 and 18.8).

At the lowest flow rate (Q =10 $\mu l \cdot min^{-1}$) the buffer exchange was slower than the periodic switching (10 s), maintaining the fluorescence intensity constant along the time (see kymograph in Fig.18.8).

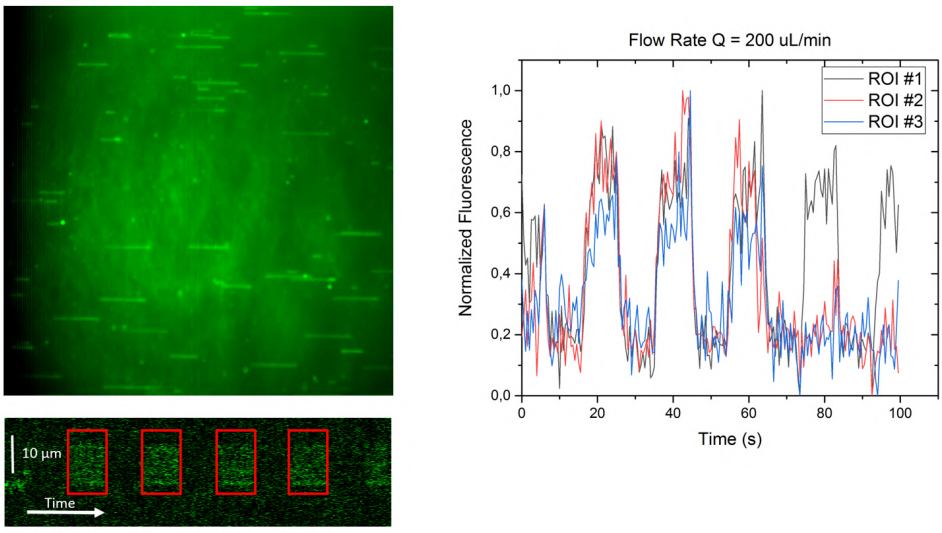

Figure 18.5.: (Top Left) Example of a fluorescence image of Sytox green stained DNA under TIRF illumination. DNA molecules were horizontally stretched on the surface due to the presence of an external flow of 200 $\mu l \cdot min^{-1}$. (Bottom Left) Representative DNA fluorescence kymograph of a stretched DNA molecule during five different Sytox cycles. (Right) Normalized fluorescence intensity on three different DNA molecules.

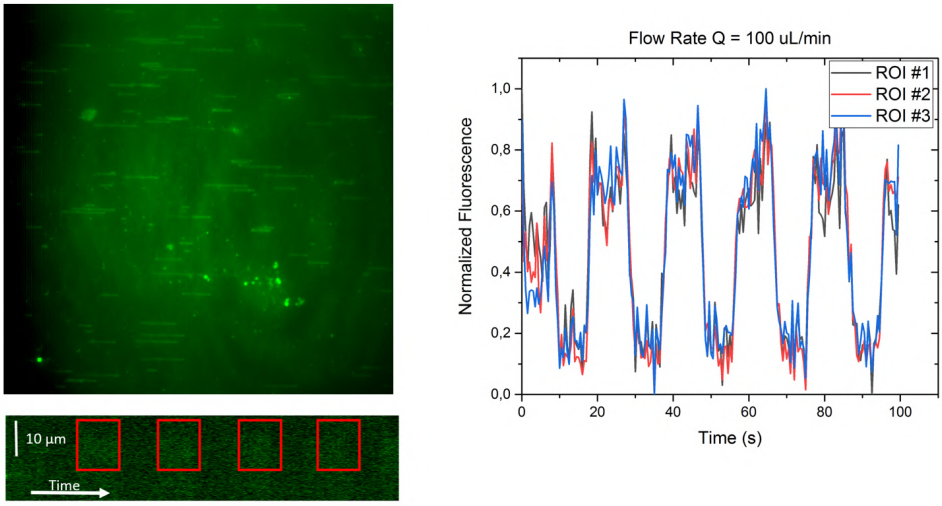

Figure 18.6.: (Top Left) Example of a fluorescence image of Sytox green stained DNA under TIRF illumination. DNA molecules were horizontally stretched on the surface due to the presence of an external flow of 100 $\mu l \cdot min^{-1}$. (Bottom Left) Representative DNA fluorescence kymograph of a stretched DNA molecule during five different Sytox cycles. (Right) Normalized fluorescence intensity on three different DNA molecules.

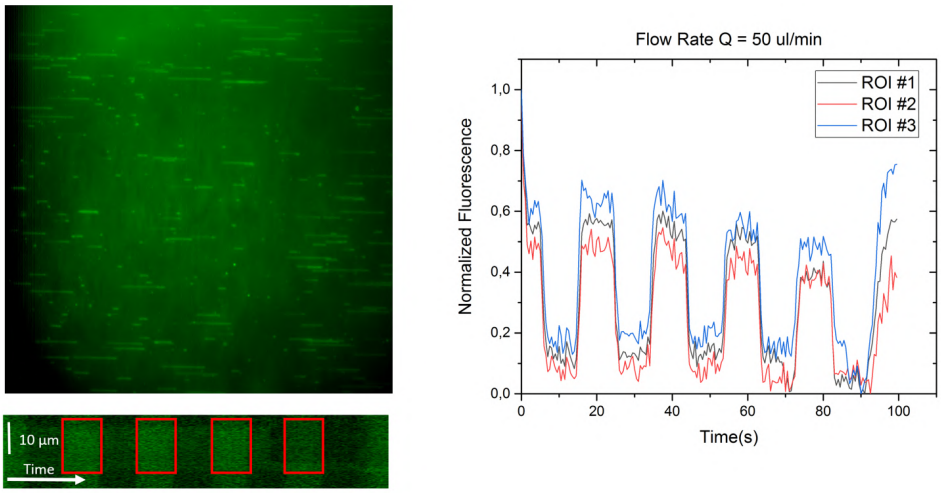

Figure 18.7.: (Top Left) Example of a fluorescence image of Sytox green stained DNA under TIRF illumination. DNA molecules were horizontally stretched on the surface due to the presence of an external flow of 50 $\mu l \cdot min^{-1}$. (Bottom Left) Representative DNA fluorescence kymograph of a stretched DNA molecule during five different Sytox cycles. (Right) Normalized fluorescence intensity on three different DNA molecules.

Moreover, it could be noted that the DNA molecules were totally extended also in the presence of the lower flow rate, as qualitatively shown in Fig.18.9.

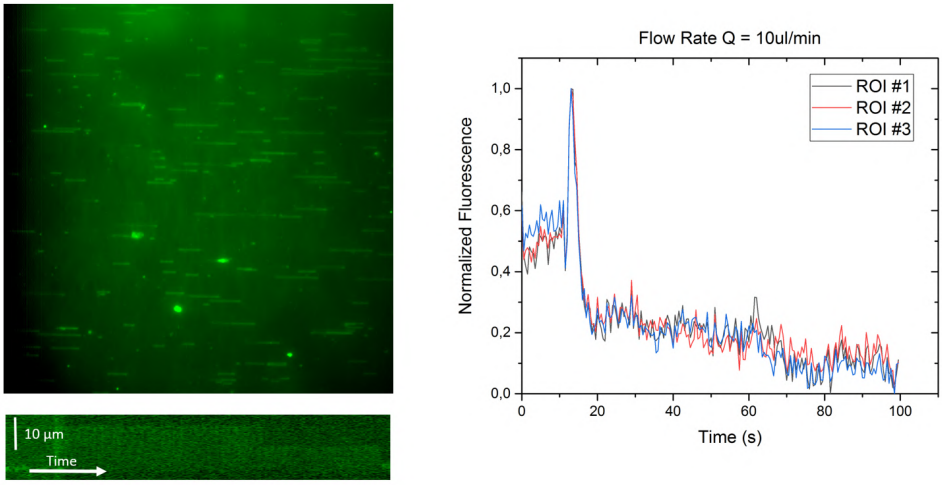

Figure 18.8.: (Top Left) Example of a fluorescence image of Sytox green stained DNA under TIRF illumination. DNA molecules were horizontally stretched on the surface due to the presence of an external flow of 10 $\mu l \cdot min^{-1}$. (Bottom Left) Representative DNA fluorescence kymograph of a stretched DNA molecule during five different Sytox cycles. (Right) Normalized fluorescence intensity on three different DNA molecules. The flow rate was not enough to achieve a switched between Sytox and buffer solution.

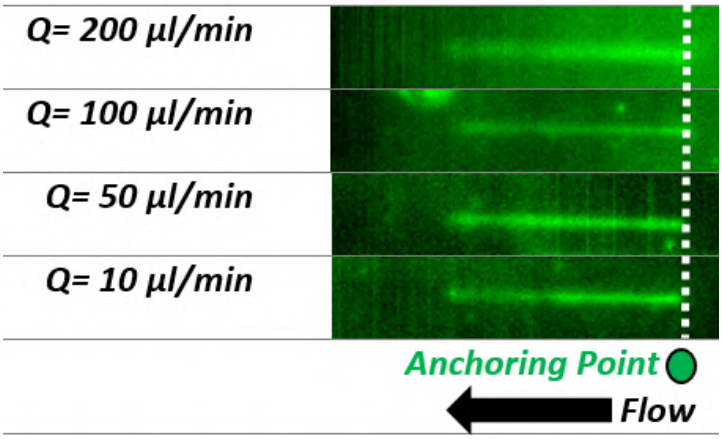

Figure 18.9.: Representative DNA molecules stretched at different flow rates. Qualitatively there is snot any difference between the lengths of the DNA molecules.

We recorded the fluorescence of several DNA molecules at different flow rates to characterize the Sytox binding to DNA. Due to the low fluorescence of Sytox when is free in solution, the fluorescence intensity of the background was constant during the entire experiments (data not shown).

We fitted our data with the Taylor-Aris model (Fig.18.10 and Tab.18.4). The fit was quite good even if the final signal was a results of both Sytox binding and dye diffusion. Additionally, our measurements included effects from photobleaching which were difficult to account for.

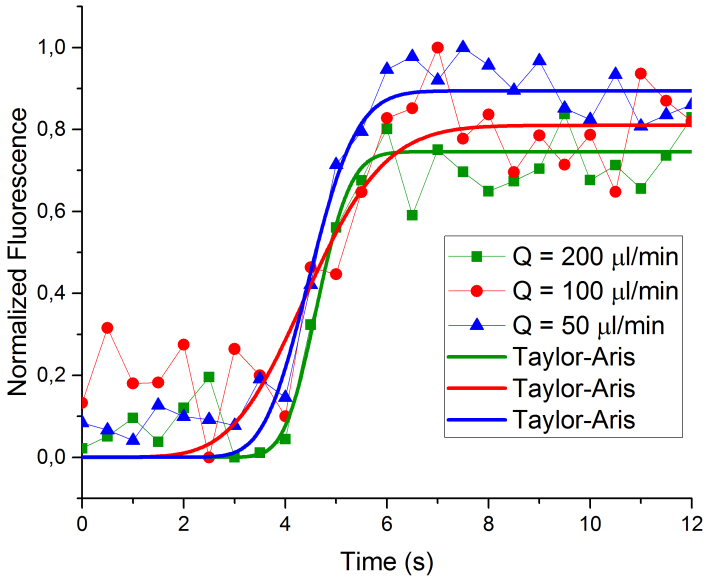

Figure 18.10.: Taylor-Aris fit on normalized fluorescence intensity at three different flow rates Q. Data shown (symbols) were averaged on 5 points. For all the flow rates considered the fluorescence saturation was reached in less than 2 s. Extracted parameters are reported in Tab.18.4.

The fluorescence intensity along the DNA molecule increased in the presence of Sytox and decreased when buffer alone was fluxed. Moreover, the dynamics of the system DNA:Sytox could be easily solved assuming that the concentration of Sytox remained constant during the injection phase. This was a reasonable assumption, there was an excess of dye compared with DNA and our measurements were significantly longer than the boundary shifting time. Then, the

Flow rate (Q) $(\mu l \cdot min^{-1})$	v (m/s)	y_0 (m)	D (m^2/s)	a (m)
200	$1.38 \cdot 10^{-4}$	$1.41 \cdot 10^{-3}$	$1.77 \cdot 10^{-10}$	$1.77 \cdot 10^{-5}$
100	$1.21 \cdot 10^{-4}$	$1.82 \cdot 10^{-3}$	$3.42 \cdot 10^{-10}$	$4.12 \cdot 10^{-5}$
50	$1.31 \cdot 10^{-5}$	$1.49 \cdot 10^{-3}$	$3.78 \cdot 10^{-10}$	$1.05 \cdot 10^{-5}$

Table 18.4.: Independent Taylor-Aris fit for the three considered flow rates (Q = 200, 100 and 50 $\mu l \cdot min^{-1}$) . 4 free parameters: y_0 (distance from inlet), v (fluid speed), a (channel height), D (diffusion constant).

final expression for the fluorescence signal of a dye binding to DNA [43, 62] could be expressed as:

$$F(t) = F_{max}(1 - e^{-k_{obs}t}) \tag{18.3}$$

where $F(t)$ stated for the fluorescence signal and the k_{obs} was the observed binding rate (obtained k_{obs} were reported in Fig.18.11).

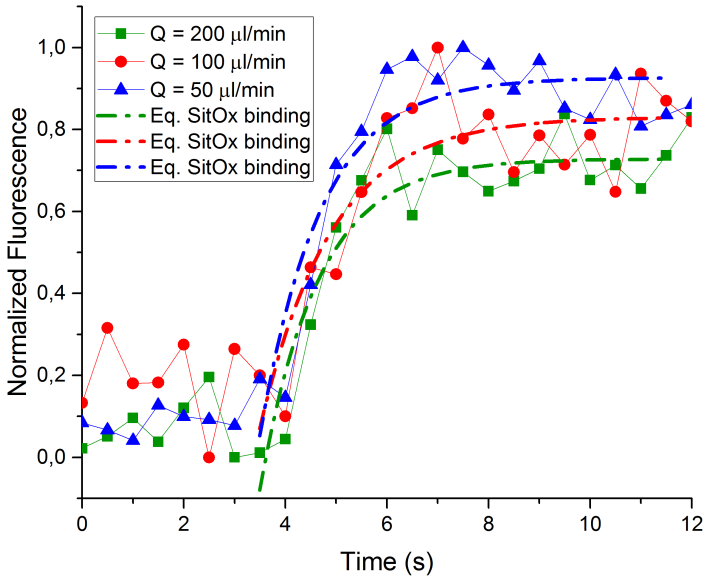

Figure 18.11.: Normalized fluorescence intensity at three different flow rates Q fitted to Eq.18.2.1 (dot-dashed lines). The observed constant k_{obs} of association of Sytox to DNA molecule are $k_{obs}(Q = 200) = 0.9 \pm 0.2s^{-1}$,$k_{obs}(Q = 100) = 0.8 \pm 0.2s^{-1}$ and $k_{obs}(Q = 50) = 0.7 \pm 0.2s^{-1}$

In a flow stretch experiment, each DNA molecule was elongated close to the surface by the external force, generating by the applied flow. To characterize the force of the flow as a function of its flow rate, we took advantage of a bead attached to one end of the DNA. In this way, we could precisely measure the extension of the tether by tracking the bead with both the standard MT CCD camera and software (Fig.18.12).

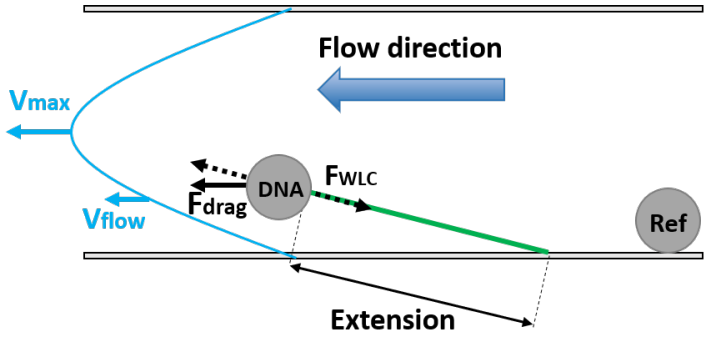

Figure 18.12.: The DNA molecule was stretched when a flow was applied. The extension and associated forces could be measured using MT CCD camera for tracking the bead position. Figure modified from [62]

Experiments were performed with 1µm beads and λ/2 DNA molecules in a 1 mm - 1 layer flow cell using both the syringes of the NeMESYS system in parallel to achieve a final flow rate Q (Q=5 - 50 $\mu l \cdot min^{-1}$). As clearly shown in Fig.18.13, the vertical DNA extension was drastically affected by the application of flow, even with a very low flow rate (at Q = 10 $\mu l \cdot min^{-1}$ the vertical extension of DNA was roughly half of the molecule length in the absence of flow).

Extension versus flow data could be correlated with the applied force using previously-taken force-extension curve performed with the vertical MT setup as described in [44, 62].

Briefly, considering a laminar flow, the bead was affected by a drag force given by Stokes' law:

$$F = 6\pi R\eta v_{flow} \tag{18.4}$$

where R is the radius bead, η the flow viscosity and v_{flow} could be estimated as $v_{flow} = kv_{max} = \frac{2 \cdot Q}{w \cdot d}k$ where Q is the flow rate, k is the ratio of fluid velocity to be taken in account an depends to the distance from the middle of the channel and d and w correspond to the channel height and width, respectively.

The viscosity η, should be corrected because the radius of the bead was comparable to the distance of the bead to the surface [44]. At $z \sim 1\mu m$ and $R = 1\mu m$ the corrected viscosity was $\eta^* = 1.6\eta$.

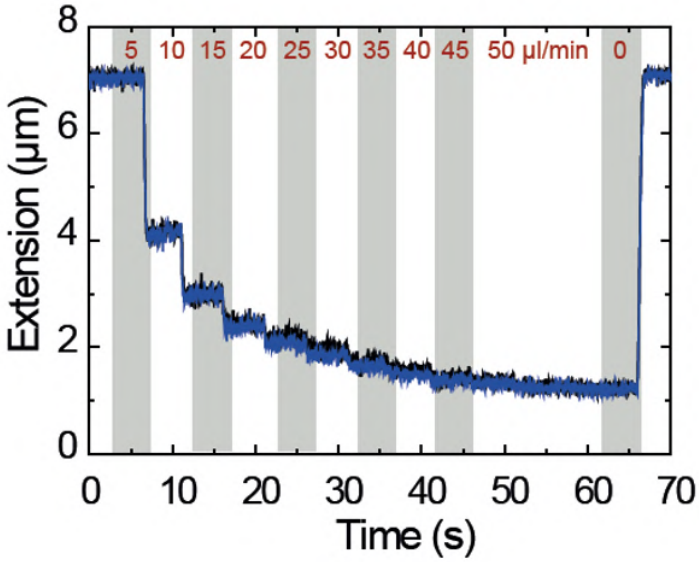

Figure 18.13.: Vertical DNA extension versus the function of the applied flow rate in narrow flow cell.

It was possible then to estimate the linear velocity of the flow in the proximity of the bead, by fitting Eq.18.5 and 18.6 to extension and force data as a function of the applied Q (Fig.18.14 and 18.15):

$$l(Q) = L\left(1 - \frac{1}{2}\sqrt{\frac{k_B T}{P \cdot 6\pi R\eta^* \frac{2\cdot Q}{w\cdot d}}}\right) \tag{18.5}$$

$$F(Q) = 6\pi R\eta^* \frac{2 \cdot Q}{w \cdot d} k \tag{18.6}$$

For our experiment, $k_B T = 4.11 pN \cdot nm$, $P = 40nm$, $R = 1\mu m$, $\eta^* = 1.6 mPa \cdot s$, $w = 1mm$ and $d = 100\mu m$.

By fitting Eq.18.5 to the extension data (Fig.18.14) we obtained a length of $\lambda/2$ DNA $L = 8.4 \pm 0.1\mu m$ and a k of $k = 1.7 \pm 0.2\%$, with a final velocity in the proximity of the surface for the bead of $v_{flow} = 1.7\% v_{max}$.

Then, we fitted the force versus flow rate data (Fig.18.15) with Eq.18.6, obtaining a value of $k = 1.8 \pm 0.2\%$, with a final velocity in the proximity of the surface for the bead of $v_{flow} = 1.8\% v_{max}$, very close to the one obtained for extension versus flow rate data.

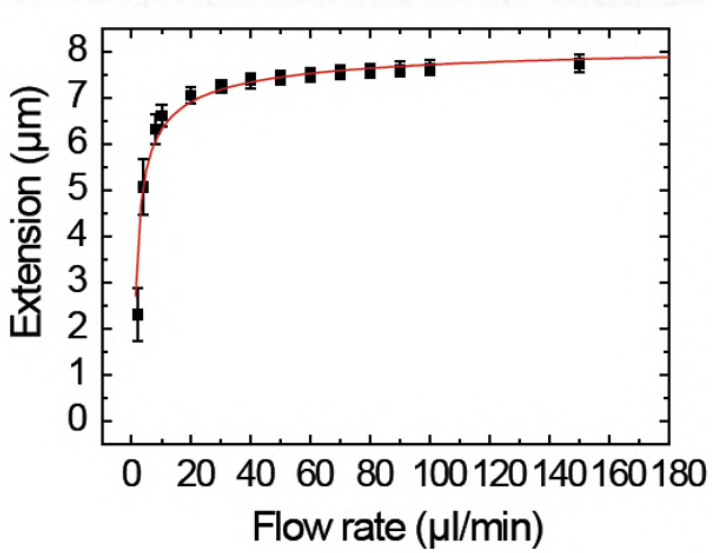

Figure 18.14.: Mean DNA extension measured as a function of flow rate in a flow stretch experiment using 1 µm beads and $\lambda/2$ DNA molecules. The solid line is the fit to Eq.18.5 with P = 40 nm as a fixed parameter, obtaining L = 8.4 µm. Error bars were the standard deviation of the mean from measurements of 9 beads.

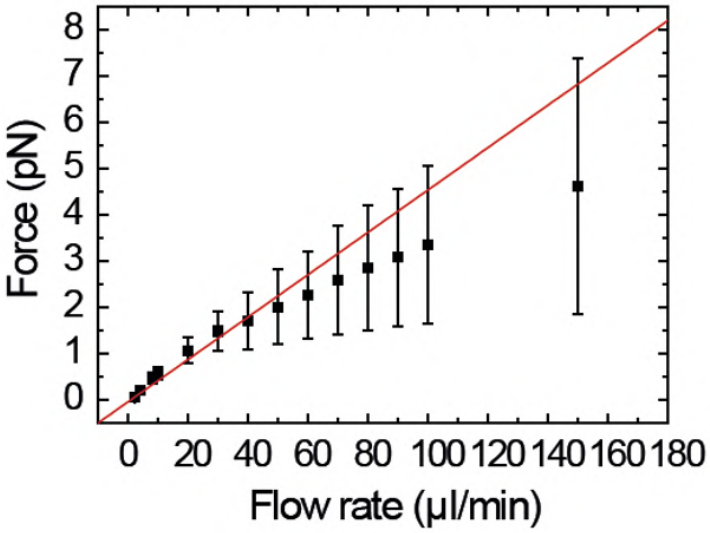

Figure 18.15.: Mean force measured as a function of flow rate. The force was determined from a calibration force-extension curve obtained from the vertical MT in the absence of flow. The force increases linearly up to a maximum value of ~ 4.5 pN, in accordance with Eq.18.6. Error bars were the standard deviation of the mean from measurements of 9 beads.

19

DISCUSSION: SIMPLE AND ROBUST FLOW STRETCHING APPARATUS FOR SINGLE MOLECULE FLUORESCENCE EXPERIMENTS

The implementation of narrower flow cells for flow stretching measurements allowed a rapid exchange of buffer without affecting the behaviour of the stretched molecules. The fabrication of the flow cells did not require any special attention and, contrary to what happened in the old design [62], the introduction of bubbles could have been avoided with a careful injection of the sample. Moreover, these flow cells were extremely versatile and could be employed in both lateral and vertical MT experiments as well as in TIRF-based essays.

The switch time t_s between two different solutions was lowered to less than 2 s in a remarkable range of applied flow rates (with less than 1 s for $Q > 100\ \mu l \cdot min^{-1}$). For the old design the t_s was approximately of 4 s at $Q = 200\ \mu l \cdot min^{-1}$ [43]: the boundary exchange of the new flow cells was at least 4 times faster than the old ones.

The Sytox-labelled DNA results reported in Section18.2 represent a good preliminary test for more sophisticated designs of DNA stretched experiments. The same experimental setup (two inlets - one outlet) could be easily employed for testing DNA:protein binding, both in the presence of fluorescently-labelled proteins and intercalating dyes, or to visualize DNA superstructure as supercoiling [53]. The next step could be the labelling of proteins with QDs to obtain a co-localized two channel fluorescent image of, for example, protein and Sytox-stained DNA.

On the other hands, these preliminary results could be employed to construct a multi-channel flow cell, as, for example, the one employed in [54], with three inlets and one outlet. By using such an approach, the flow could be applied transversely to a DNA molecule anchored to the surface by both the ends. This is an essential test in the frame of proteins interacting with DNA, like condensin and SMC complexes.

Finally, by monitoring single DNA extension and using individual force extension curves, we were able to estimate the forces exerted on a bead in flow-stretch experiments. We showed that measured forces were sensibly higher than the one previously reported with the old flow cell design [44, 62] going from less than 1.5 pN to 4.5 pN. However, the calculated forces were still more dispersive than those

obtained in both vertical and lateral MT, probably due to the presence of a not perfect laminar flow in the vicinity of the surface.

Our custom-designed flow cells are very cheap, as they use consumables usually employed in laboratory. The surface is made by a glass coverslip, which can be differently washed with chemicals and functionalised in several ways. With respect to other flow cells designs, such as the PDMS microfluidic device described in [59], our approach is faster and do not require any dedicated or specific instrument. On the other hands, the use of paraffin wax gasket limit the minimum size of the cross-section of the channel up to $\sim$ 1 mm, limiting also the applied flow rate. Furthermore, with respect to similar works performed by Dekker and colleagues [53, 54, 255], which used biotinylated-PEG, we employed the standard polystyrene - antidigoxigenin - BSA functionalisation, obtaining good results both in terms of signal-to-noise ratio during DNA-based experiments and, simultaneously, preventing the attachment of DNA to coverslip's surface. This latter functionalisation is particularly suitable in protein:DNA experiments. Indeed, a large number of the proteins employed in single molecule assays present a Streptavidin-Tag useful both for the proteins purification and for latter QDot-labelling. Avoiding free biotin sites in the surface is particularly crucial for certainly distinguish between protein:surface and protein:DNA interactions.

19.1 CONCLUSION

In conclusion, we demonstrated the crucial role of both the shape and dimension of flow cells for single molecule experiments and we proposed a new flow cell design which allowed, in the proximity of the inlet, to achieve a fast boundary exchange and, if necessary, to apply an external force comparable to the typical range of MT-based setups. This method was very easy to implement and it was extremely cheaper than commercially available products, as well as home-made PDMS microfluidic devices employed in other studies [59]. The reduction of the volume was also essential to limit the quantity of solutions, ligands and sample necessary to perform the experiment. Moreover, this flow cell could be completely customizable: the glass surface could be cleaned and functionalized in several different ways (coated with antibodies, PEG or silane, evaporated with gold or other metals). Furthermore, by spanning among all the flow cells employed, it was possible to finely chose the design which better achieved the requirements of the single experiment. For example, MT experiments requiring a fast buffer exchange and low perturbation could take advantage of wider flow cell, while experiment with a very expensive (or difficult to obtain) specimen would prefer the use of a flow cell with a volume as small as possible.

Part VI

SUMMARY

CONCLUSIONS

In the present Thesis, I employed single molecule force spectroscopy (SMFS) for studying biological macromolecules in 4 different main projects. In particular, I studied the (1) conformational changes of the α-synuclein (AS) protein in the presence of ligands or point mutations, (2) the nanomechanical behaviour of a DNA analogue, (3) the condensation of DNA mediated by the condensin protein and (4) a technical implementation on Flow-Stretching (F-S-TIRF) setup to achieve high speed buffer exchange.

The conclusions for these projects are briefly summarized below.

1. I have directly monitored the conformational changes of the AS protein in solution by SMFS, focusing on their changes in the presence of ligands (DA and EGCG) and of point mutations. The SMFS description of AS conformational ensemble in solution detected differently structured components, in agreement with the different degrees of compactness suggested by native mass spectrometry. I also discussed the presence of a secondary structure by using CD and FTIR techniques. Our results might provide valuable constraints for computational simulations of intrinsically disordered proteins (IDPs) conformational ensembles in the presence or absence of ligands and new insight into the Parkinson's Disease (PD) development linked to gene mutations.

2. I have shown, through AFM imaging, MT-based mechanical characterization, and CD spectroscopy, that DAP-substitution induces DNA molecules to adopt a non-canonical form when in relaxed configuration and it resists to flexural bending at low-to-moderate tension (<10 pN), yet it also decreases the overstretching threshold from roughly 60 pN to 52 pN. From an energetic point of view, the DAP extra amino group penalizes entropic bends but, it also raises the conformational free energy of the B-form helix, facilitating axial extension and initiating the B-to-S transition at a lower tension.

3. I have directly visualized the interactions of condensin complexes with DNA, at different forces and with different DNA configurations. I have implemented two algorithms to individuate steps in single molecule traces. Our results have revealed that the condensin compacts the DNA in a multistep mechanism, with a step size depending on the tension exerted on the molecule.

4. I have achieved a multistream laminar flow by the implementation of a new flow cell design which allowed to achieve a fast boundary exchange in the proximity of the inlet. I tested this flow cell in terms of both buffer exchange, with a resolution 0.5 s that minimized diffusion, and force, reaching a maximum value of 4.5 pN. I have employed the setup also for the stretching of fluorescently-labelled DNA molecules.

GENERAL PERSPECTIVE

The SMFS, often coupled with other techniques, i.e. fluorescence, native MS and biophysical standard methods, certainly improves the knowledge of several different biological processes. The SMFS approach permits a very precise control of the applied forces, as well as it provides unique features, such as the possibility of impose a torque for MT or the great versatility in both imaging and force for the AFM apparatus. The combination of a an high sensitive force spectroscopy methods with fluorescence would therefore have a striking impact in the single-molecule technology.

Additionally, great interest is raising in both the fields of IDP characterisations and SMC mediated DNA compaction mechanism. Recent studies have proposed that the unstructured form of the IDPs may be the final point of evolutionary process. For this reason, it would be very interesting to provide new information on the plasticity conformations of this family of proteins as well as to fully depict their role in the insurgence of several diseases. On the other side, it would be interesting to fully understand the role of SMC. These questions have been addressed by both *in vivo* and *in vitro* experiments, but the detailed mechanics remain elusive. A complete reconstitution of DNA compaction by SMC complexes at the single molecule level is required to fully decipher the genome organization in bacteria.

PUBLISHED PAPERS AND OTHER WORKS

A

AFM-BASED SINGLE MOLECULE IMAGING AND COLLAGEN CHARACTERIZATION

During my PhD program I had the opportunity to collaborate with different and quite interdisciplinary groups. In this chapter I will present some examples of these AFM imaging side-projects, which have been published (Section A.1.1 and A.3,[3, 4]) or are into the review process (Section A.1.2, A.2).

A.1 β-AMYLOID PROTEIN

Alzheimer's disease (AD) is the most common type of dementia and affects tens of millions of over-6-aged people worldwide. One of the most proposed hypothesis, *the amyloid hypothesis* [256, 257] proposes that β-amyloid peptide (Aβ), the main component of senile plaques, is the key player in AD pathogenesis. Aβ monomers, derived from the proteolytic cleavage of the larger glycoprotein amyloid precursor protein, if not efficiently cleared from the brain can aggregate into different assemblies, which can then form regular fibrils and plaques that are involved in the insurgence of the disease. The aggregative behavior of Aβ peptide has been extensively analyzed since many years. In the next sections the effects on the Aβ aggregation of human lipoproteins (Sec A.1.1) or the synthetic oligopeptide mApoE absorbed on gold nanoparticles (Sec A.1.2) has been studied both by AFM-based imaging and fluorescence essay. An *in vitro* approach for monitoring aggregation/disaggregation of the Aβ peptide has been used as briefly described in the following paragraph. For more details see [4].

Preparation of Aβ samples and AFM imaging

In order to obtain monomer-enriched $A\beta_{1-42}$ preparations, the $A\beta_{1-42}$ peptide (1 mg/ml) was solubilized in 1,1,3,3,3-hexafluoro-2-propanol (HFIP; Sigma–Aldrich) dried and then resuspended in DMSO at a concentration of 5 mM and bath sonicated for 10 min.

To follow the $A\beta_{1-42}$ aggregation, the 5 mM $A\beta_{1-42}$ sample in DMSO was divided in aliquots immediately after sonication. Each $A\beta_{1-42}$ aliquot was diluted to a final concentration of 100 μM in 10

mM HCl aqueous solution and was incubated at 37°C. If requested, the sample was incubated in the presence of mApoE NPs. At selected incubation times 1 µl of each aliquot was withdrawn from the sample and characterized by AFM.

To follow the $A\beta_{1-42}$ disaggregation, the started sample was a fibril-enriched preparation: immediately after sonication the 5mM $A\beta$ sample in DMSO was diluted to 220 µM in 10 mM HCl and incubated at 37°C for 72 h. $A\beta_{1-42}$ fibrils preformed *in vitro* were diluted to 100 µM and incubated in the presence of either mApoE coated NPs or lipoproteins at 37°C in NaCl 10mM (pH 7). After different times of incubation 1 µL of each sample aliquot was characterized by AFM.

For AFM imaging 1 µL of each sample aliquot was diluted 1:10 in HCl 10 mM and incubated for 5 min on a freshly cleaved mica substrate. After incubation, samples were rinsed with Milli-Q water and dried in a gentle stream of nitrogen. Measurements were performed using a Nanowizard II (JPK Instruments, Berlin) scanning probe microscope operating in tapping mode in air (Section 1.2.1). RTESP-300 (Bruker, USA) cantilevers were used with a nominal force constant of 40 N/m, a resonance frequency of 300 kHz, and a nominal tip radius of 8 nm. The AFM images were analyzed using the commercial JPK image processing software and a customized image-analysis software (Matlab, MathWorks Inc). All data in this study were verified by sampling a wide range of areas over the samples surfaces.

A.1.1 β-*amyloid protein interaction with LipoProteins*

In AD progression, most evidence suggests a protective role of high-density lipoprotein (HDL) and its major apolipoprotein apoA-I. The 28 kDa Apolipoprotein A-I (apoA-I) is known to be involved in the generation and metabolism of HDL [258]: the protein is synthesized and lipidated with cholesterol and phospholipids to create different subclasses of plasma HDL particles, including discoidal ones. Subsequently, a second lipidation step is required for maturation of nascent HDL into mature spherical lipid-rich HDL [259]. Therefore, in humans, HDL consist of heterogeneous subclasses, characterized according to charge, size, density, protein, and/or lipid composition. HDL are involved in removing excess cholesterol or other protective effects, including anti-oxidative, anti-inflammatory, anti-apoptotic, and anti-infective actions [260]. Dysregulated HDL metabolism has also been linked to brain disorders: a decrease in plasma levels of HDL and/or apoA-I are risk factors for memory decline and neurodegenerative diseases, including AD [261]. In this study, the ability of apoA-I in different lipidation states affect Aβ efflux across the BBB or the Aβ fibrils disaggregation has been assessed.

Characterization of HDL proteins

First, a characterization of the two different HDL (both discoidal and speherical) has been done by AFM imaging (Fig.A.1). Both HDL subclasses have homogeneous size distribution and morphology. The height statistical distributions of AFM imaging allowed discrimination between the spherical and discoidal shapes (Fig.A.1.C) and the average dimensions of the two HDL subclasses were comparable (spherical apoA-I-HDL diameter $2R_S$ 12.7 nm, discoidal apoA-I-HDL size 12.9 nm)

Considering the intrinsic tip convolution (finite size and specific geometry of the AFM cantilever tip) and the relatively small dimensions of the samples, the aspect of discoidal and spherical HDL cannot be clearly distinguished by AFM imaging. However, these image limitations can be overcome by considering the shape of the height statistical distributions. Indeed, as result of AFM imaging procedure it was possible to associate to each planar coordinate of the image (x, y), the quantity z which represents the measured height at the spatial coordinates. We analysed the statistical distribution versus the height for bot discoidal and spherical HDL, revealing a clear difference in the two distributions (Fig.A.1). For discoidal HDL (Fig.A.1, in green), the distribution of the height values ranging between 4 nm and 25 nm was significantly different from the height distribution of spherical HDL (Fig.A.1, in red). Moreover, we decided to study only values > 4 nm in order to eliminate the spurious substrate contributions (i.e. irregularity on the surface). To quantify the resulting outcomes, we considered the theoretical height distribution for a sphere and fit the normalized spherical HDL height histograms with the formula theoretically predicted for a sphere:

$$P(h) = n_0 \left(2\pi \sqrt{R_S^2 - (h - R_S)^2} \right)$$

Where n_0 is the normalization factor and R_S is the mean radius of the sphere. We obtained a $2R_S$=12.7 nm with a coefficient of determination R^2=0.96.

For discoidal HDL (Fig. A.1, in green) the theoretical height distribution can be considered a δ-function centred on the disks average height h_D. The height distribution is influenced by the tip convolution, the polydispersity of the sample and the possible different deposition orientation of the disks. Since the data cannot be fit with the δ-function, the mean disks height is considered the maximum height of the distribution. By this procedure, the average height value for discoidal HDL was 12.9 nm.

Characterization of Aβ Fibrillation

The aggregation process of Aβ from monomers to fibrils was followed and characterized by AFM imaging during two days of aggregation.

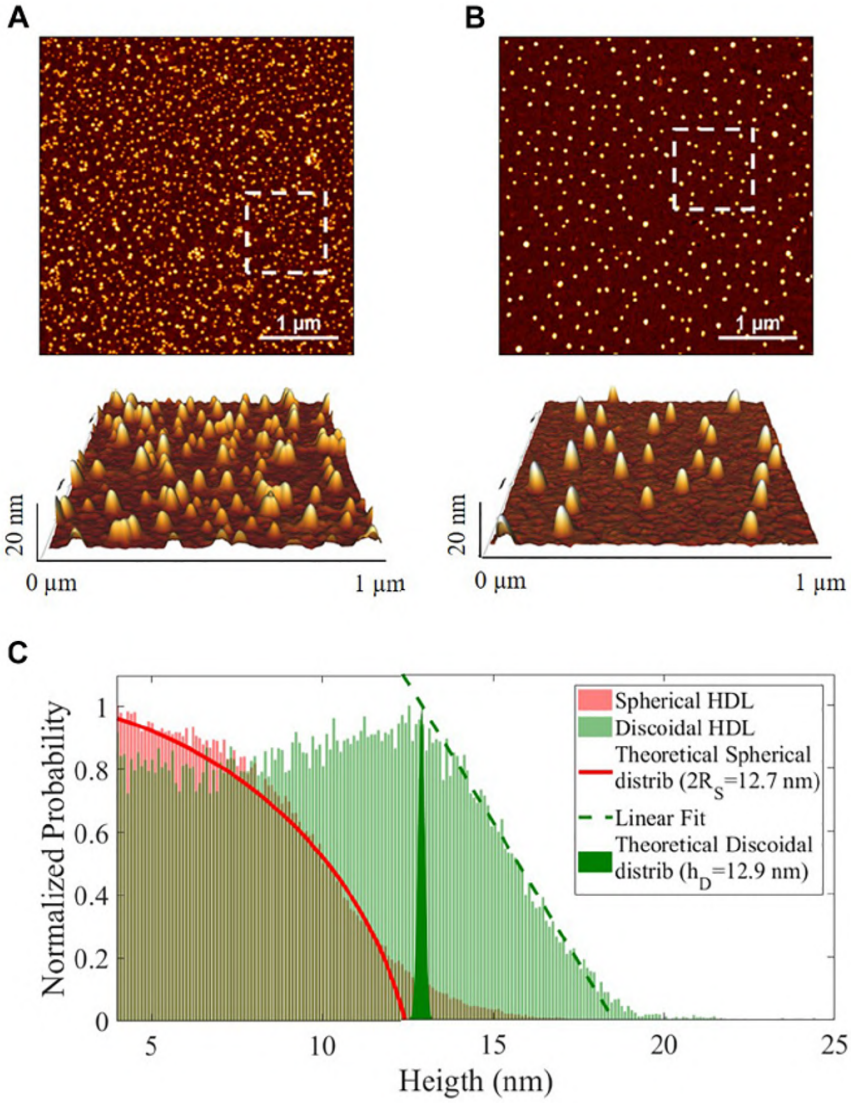

Figure A.1.: **Representative images of the morphology of HDL subclasses obtained by AFM imaging**. A) Discoidal and B) spherical HDL on APTES functionalized mica (4 x 4 μm^2, 1024 x 1024 pixel, Z-scale 20 nm), with 3-dimensional projections of a 1 x 1 μm^2 region (white squares in the images). C) Normalized height statistical distributions of spherical (red) and discoidal (green) HDL in the range 4-25 nm. The red fit is the theoretical distribution for a sphere ($2R_S = 12.7nm$), the green filled area represents the δ-function ($h_D = 12.9nm$) and the green dotted line is the linear regression for the final part of the distribution. Figure taken from [4].

Several images were acquired at the beginning of the fibrillation (t = 0 h) to confirm that the starting point had no aggregates, and at successive fixed times (t = 4, 8, 24, and 48 h) (Fig.A.2). The results showed that the aggregation process was characterized by progressive formation of unbranched fibrils of constant diameter and increasing length [262]. The process of fibril growth can be quantitatively evaluated by considering the time evolution of the number of pixels above a fixed height threshold in the AFM images of fibril morphology (Fig.A.4 and A.5). The percentage of pixels above a certain threshold increases with fibril extension and density, and it can be considered as a quantitative index of the aggregation process. Threshold pixel percentages over time show that there is continuous Aβ fibril growth in length up to 48 h.

Effect of apoA-I Lipidation on Preformed Aβ Fibrils

The effect of apoA-I lipidation on the disaggregation of preformed Aβ fibrils was assessed by AFM and thioflavine T (ThT) assay (Fig A.3). Aβ fibrils were incubated with apoA-I in different lipidation states for up to 24 h and changes in the morphology of fibrils were followed by AFM imaging (Fig.A.3). The results showed that, starting from mature Aβ fibrils of comparable length, the incubation with spherical apoA-I HDL did not induce significant changes in fibril morphology and concentration compared to Aβ fibrils alone. On the contrary, incubation with both the apoA-I-HDL plasma pool (consisting in a mixture of spherical and discoidal HDL) and discoidal apoA-I-HDL induced a strong time-dependent reduction of fibril concentration and extension. The percentage of pixels above the 1.5 nm threshold normalized with respect to the starting point (value at t = 0 h) was reported for each sample at different times to obtain a semi-quantitative analysis of AFM images. The results demonstrated the superior capability (1.6-fold increase) of discoidal apoAI- HDL in disassembling preformed Aβ fibrils compared to spherical apoA-I-HDL. This was confirmed by ThT assay, where a strong, and rapid reduction of the β-sheet content of fibrils was detected for discoidal HDL.

From imaging to numbers: semi-quantification of fibrillation

By applying a threshold to every AFM image, it was possible to measure the number of pixels whose height was above this threshold (white pixels). The number of white pixels was directly proportional to the total quantity of Aβ aggregated in fibrils in the field of view (Fig. A.2 and A.4).

In the Fig A.5, the histogram of the height distribution of a representative AFM image at a certain step of the aggregation process was shown. Two peaks were distinguishable: the first higher peak

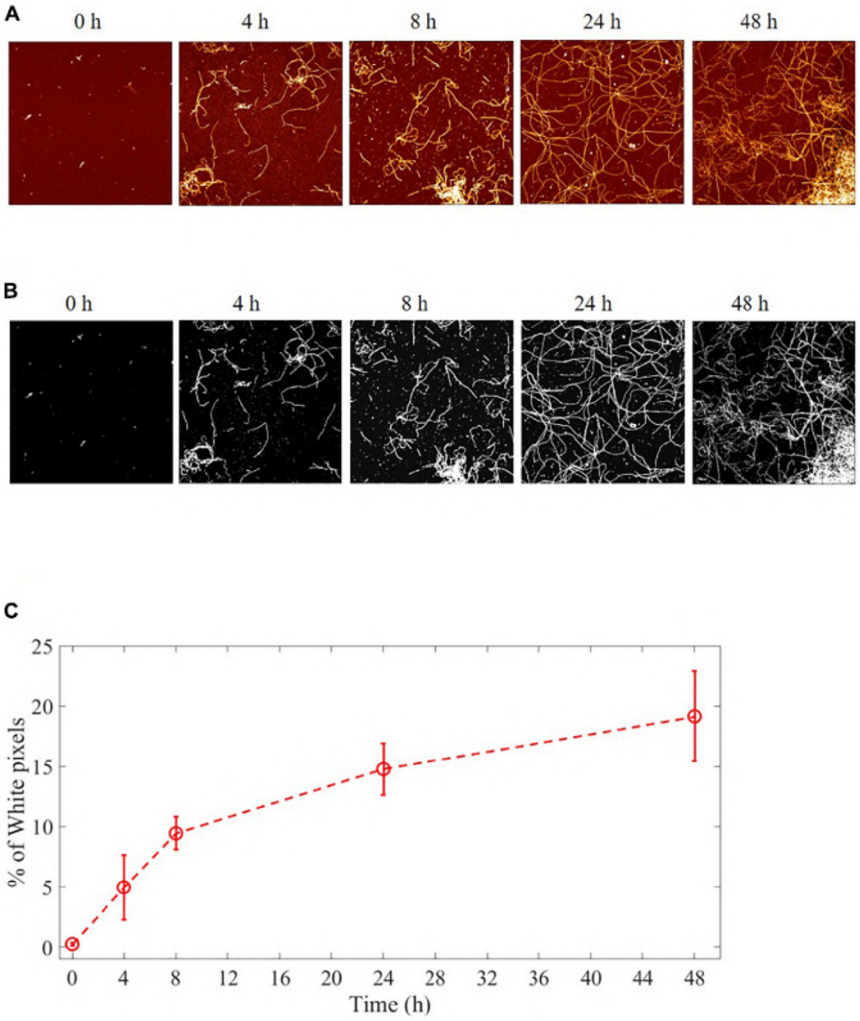

Figure A.2.: **Aggregation process of Aβ from monomers to fibrils studied by AFM.** Representative images of Aβ at different fibrillation stages (t = 0, 4, 8, 24, and 48 h) of incubation at 37 °C. A) AFM images: 4 x 4 μm^2, 1024 x 1024 pixel, Z-scale 10 nm. B) A fixed height threshold (here 1.5 nm) is applied to the AFM images in order to quantify the number of pixels above this threshold, expressed in terms of percentage with respect the total number of pixels. The number of such pixels is proportional to the total length of the fibril, i.e. to the sum of the lengths of all the deposited fibrils. In this way it is possible to quantify the fibril growth. C) Quantification of the fibrillation. Percentage of pixel above a height threshold (1.5nm) as obtained from AFM images, plotted as a function of the incubation time at 37 °C. Figure taken from [4].

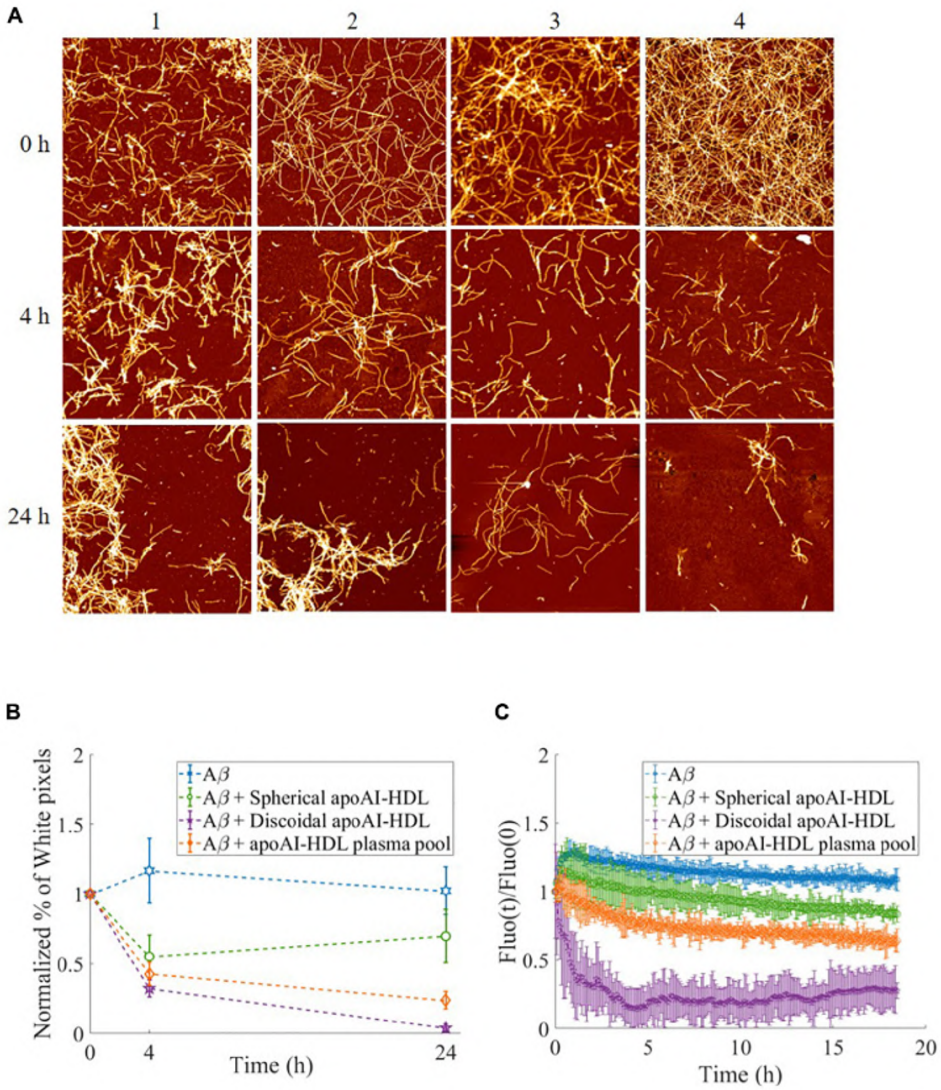

Figure A.3.: **Disaggregation of preformed Aβ fibrils in the presence of HDL.** A) Representative AFM images of Aβ fibrils over time, incubated at 37 °C either alone (column 1) or with different HDL subclasses: spherical HDL (column 2), total HDL plasma pool (column 3), discoidal HDL (column 4). 4 x 4 μm^2, 1024 x 1024 pixel, Z-scale 10 nm. (B) The normalized percentage of pixels with a height above a threshold of 1.5nm (white pixel percentage) is reported for Aβ in the presence of the different HDL subclasses at different incubation times. Values are the average of pixels higher than the threshold over several images acquired on the same sample. Error bars represent SD. Each sample is normalized to its respective starting point (value at t = 0 h). (C) Thioflavine-T fluorescence as a function of time in samples containing 2 mM Aβ fibrils alone (blue dotted) or incubated with spherical HDL (green dotted), total HDL pool (orange dotted), or discoidal HDL (purple dotted). The intensities were normalized to the respective zero-time intensity. Figure taken from [4].

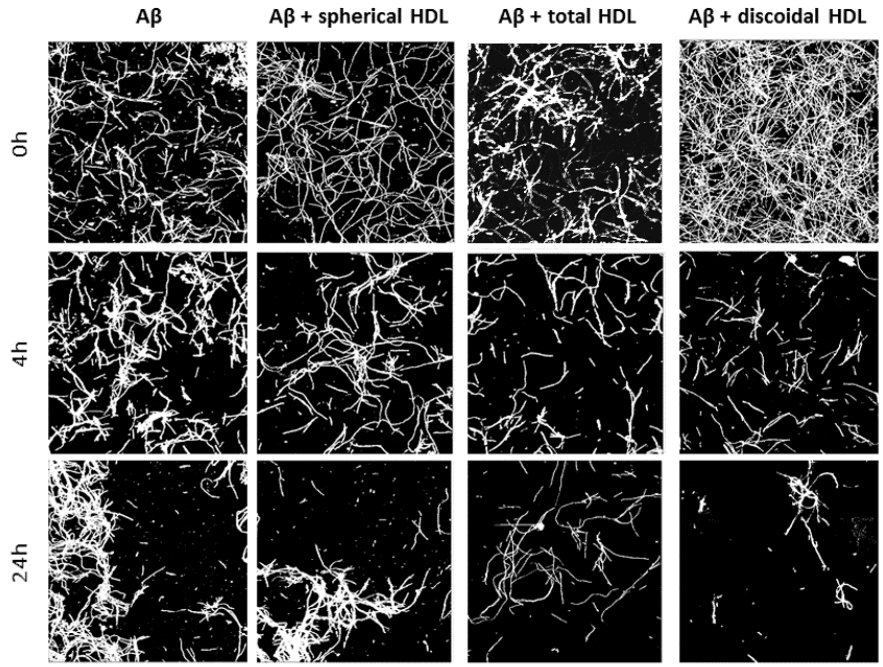

Figure A.4.: A fixed height threshold (1.5 nm) is applied to the representative AFM images Fig A.3. This process allows quantification of the percentage of white pixels (pixel above a certain threshold), which is correlated to the fibrils crowding (total length and number).

centered on h = 0 nm is due to the statistical height distribution of the mica on which the fibrils are deposited, and the other lower peak centered around h = 3-4 nm is due to the presence of fibrils on the mica. Given the partial superposition of the two peaks, the choice of a height threshold could be somewhat arbitrary. Therefore, the consequences of a particular threshold selection on the time evolution of white pixels (i.e. pixels above that threshold) for the *Aβ + Discoidal HDL* column of Fig A.3 was examined. The choice of a threshold between 0.5 nm and 2 nm is basically irrelevant as the white pixel percentage is independent of the threshold choice.

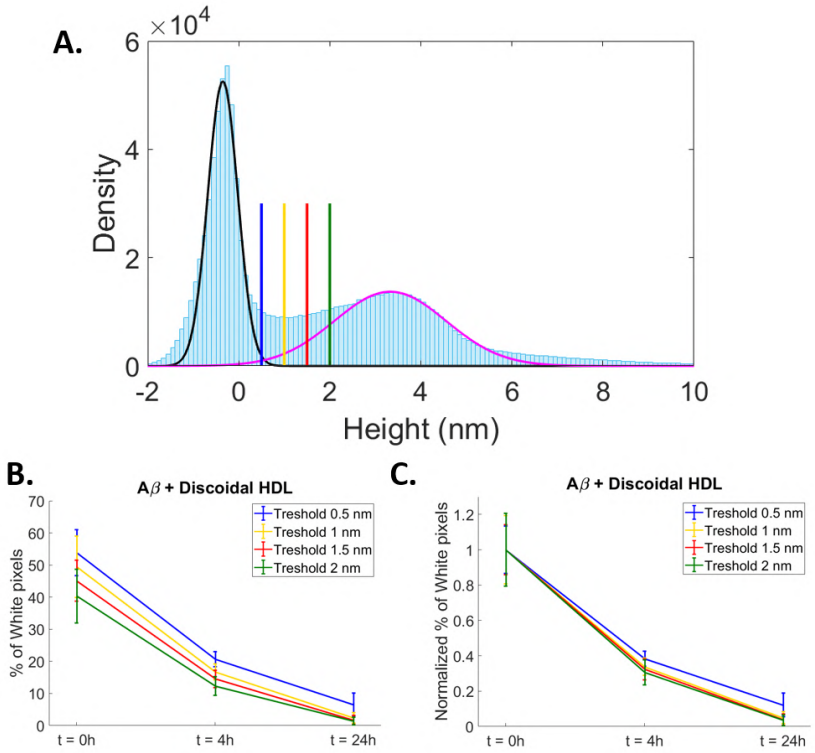

Figure A.5.: A) Representative height histogram of an AFM image of fibrils for Aβ incubated with discoidal HDL (column 4 in Fig.A.3). Two peaks are visible, one for the mica surface (black Gaussian fit), and the other corresponding to the fibrils (magenta Gaussian fit). Four possible height threshold values are indicated: 0.5 nm (blue), 1 nm (yellow), 1.5 nm (red), 2 nm (green). B) Percentage of pixels above a fixed height threshold (white pixels) as a function of the incubation time evaluated for the 4 different thresholds. C) Normalized percentage of white pixels above a fixed height threshold as a function of the incubation time. The normalization is calculated with respect to the t=0h sample. Figure taken from [4].

A.1.2 β-amyloid protein interaction with mApoE-functionalized AuNPs

Several compounds or drugs, have been designed for the purpose to redirect or stop Aβ aggregation, including peptides: for instance, the trideca-peptide CWG-LRKLRKRLLR (mApoE), derived from the receptor binding sequence of apolipoprotein E, shown to be effective in controlling Aβ aggregation. In this study the capability of mApoE functionalised gold nanoparticles (AuNPs) in the alteration on fibril aggregation and disaggregation was studied. Results showed that the performance of the mApoE peptide, in preventing Aβ aggregation and destroying Aβ preformed fibrils, is amplified when it is clustered on the AuNP surface with respect when it is free in solution. These data suggest that the clustering of Aβ-targeting molecules could be crucial in controlling both processes.

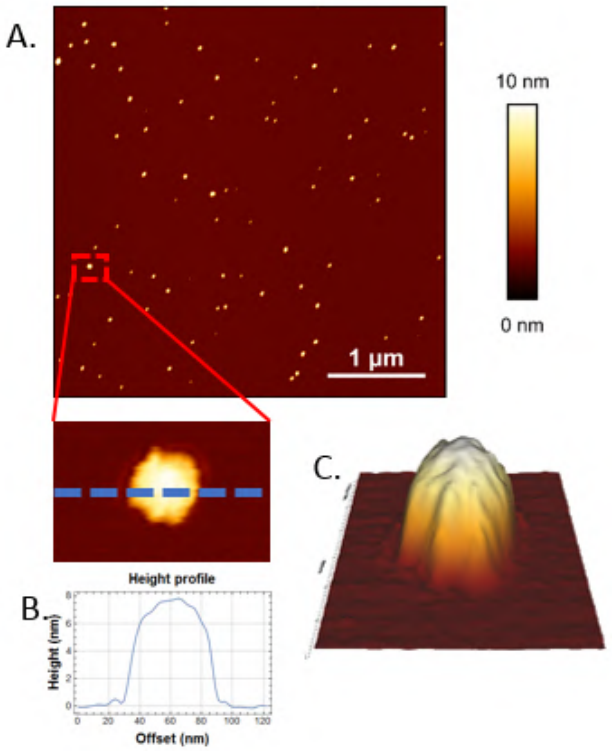

Figure A.6.: **AFM characterization of non-functionalized AuNPs on mica substrate**. A) Representative AFM image of gold NPs deposited on mica. B) Cross section of the NP highlited in the inset (average NPs radius = 5 ± 2 nm). C) A 3-dimensional magnification of one NP.

Imaging analysis and fibril recognition

We employed the procedure proposed in [263] for the fibrils recognition. This post-processing protocol had the capability to recognise

fibrils (and measure both fibril length and width distributions) also in the presence of a noisy background. In this protocol, each image is analyzed in the context of an orientation Map obtained through a five-step process: i) fibril smoothing by coherence-enhancing anisotropic diffusion filtering; ii) contrast enhancement by top hat filtering; iii) binarization by thresholding (classifying which pixels belong to fibers and which do not); iv) skeletonization of the fibers to single-pixel; v) orientation mapping from the result of diffusion filtering. Even if this method can be partially affected by an overestimation of the width of the fibrils, it results to be a quite reliable semi-quantitative method for characterizing the aggregation and disaggregation processes.

Gold Nanoparticles characterisation

The size and polydispersity of AuNPs were checked by Dynamic Light Scattering (DLS). Every sample resulted monodispersed with an average radius of 22 $\pm$ 7 nm. The NPs dimension was also verified by AFM imaging (Nanowizard II, JPK, Berlin). In Fig A.6A, a representative image of NP deposited on mica substrate was reported (4x4 μm^2, 2048 x 2048 pixel, Z-scale 5nm). From the height distribution analysis, an average particle radius of 5 $\pm$ 2 nm, value compatible with the one obtained from previously obtained TEM images [264], while the DLS measurements are particularly prone to an over-estimation of the hydrodynamic radius.

Effect of mApoE-NPs on Aβ peptide aggregation

Focusing on the first column of Fig.A, in which images of bare $A\beta_{1-42}$ at zero time or after 24 h and 48 h of incubation are reported, a progressive increase in time of the deposited fibrils number and length could be observed. The presence of non-functionalized NP (Fig.A, second column), appeared to favor the aggregation process, since after 48 h the fibrils were even more abundant than in the corresponding bare $A\beta_{1-42}$ frame. Conversely, the presence of both free mApoE and AuNP functionalized with mApoE clearly hindered the aggregation process. Apparently, the AuNP functionalized with mApoE were more efficient than free mApoE in decreasing the $A\beta_{1-42}$ aggregation rate. In order to corroborate the above described qualitative observations about the stimulated or inhibited fibril aggregation process from a quantitative standpoint, an automatic algorithm allowing to analyze the images and to extract length, width and number of deposited fibrils in each sample was applied. Confirmed also by visual inspection of the images, the procedure eliminated the noise, yet preserving the elongated objects. Disregarding spherical objects, but not elongated aggregates, the algorithm recognized the fibrils in the AFM images and reports their width and length distributions, from which it was possible to extract the percentage of

surface covered by fibrils, confirming the same trend of both visual inspection of AFM data (Fig.B)and of the FITC-Fluorescence emission (Fig.C). Notably, the FITC-Aβ_{1-42} fluorescence assay allows to unravel the initial (pre-fibrillation) steps of amyloid aggregation, which are blind to AFM analysis, while it is unfit to probe fibrillation. Thus, the combination of the two methods offers a complete panorama of the aggregation dynamics.

A.1.3 *Disaggregation effect of mApoE-NPs on Aβ preformed fibrils*

Fig.A.9A shows some representative AFM images obtained from a solution of preformed Aβ fibrils in the presence of free AuNP (second column), free mApoE (third column)or and AuNP functionalized with mApoE (fourth column). Fig.A.10B shows the percentage of deposited fibrils, normalized with respect to the initial situation. Both visual inspection of Fig. A.9A, A.9B and the quantitative analyses reported in Fig.A.10A and A.10B suggested that the amount of deposited fibrils was essentially constant in time for bare Aβ, while the presence of non functionalized or mApoE functionalized AuNP causes a progressive disaggregation of the preformed fibrils, which was more effective in the latter case. Notably, even bare NP display a moderate but non-negligible disaggregating potential on preformed fibrils. Interestingly, the mApoE-functionalized NP induced the steeper reduction in the mean number of fibrils, and were apparently more effective in reducing their average length. The above results were qualitatively confirmed by those obtained by using the ThT-assay which are reported in Fig.A.10C.

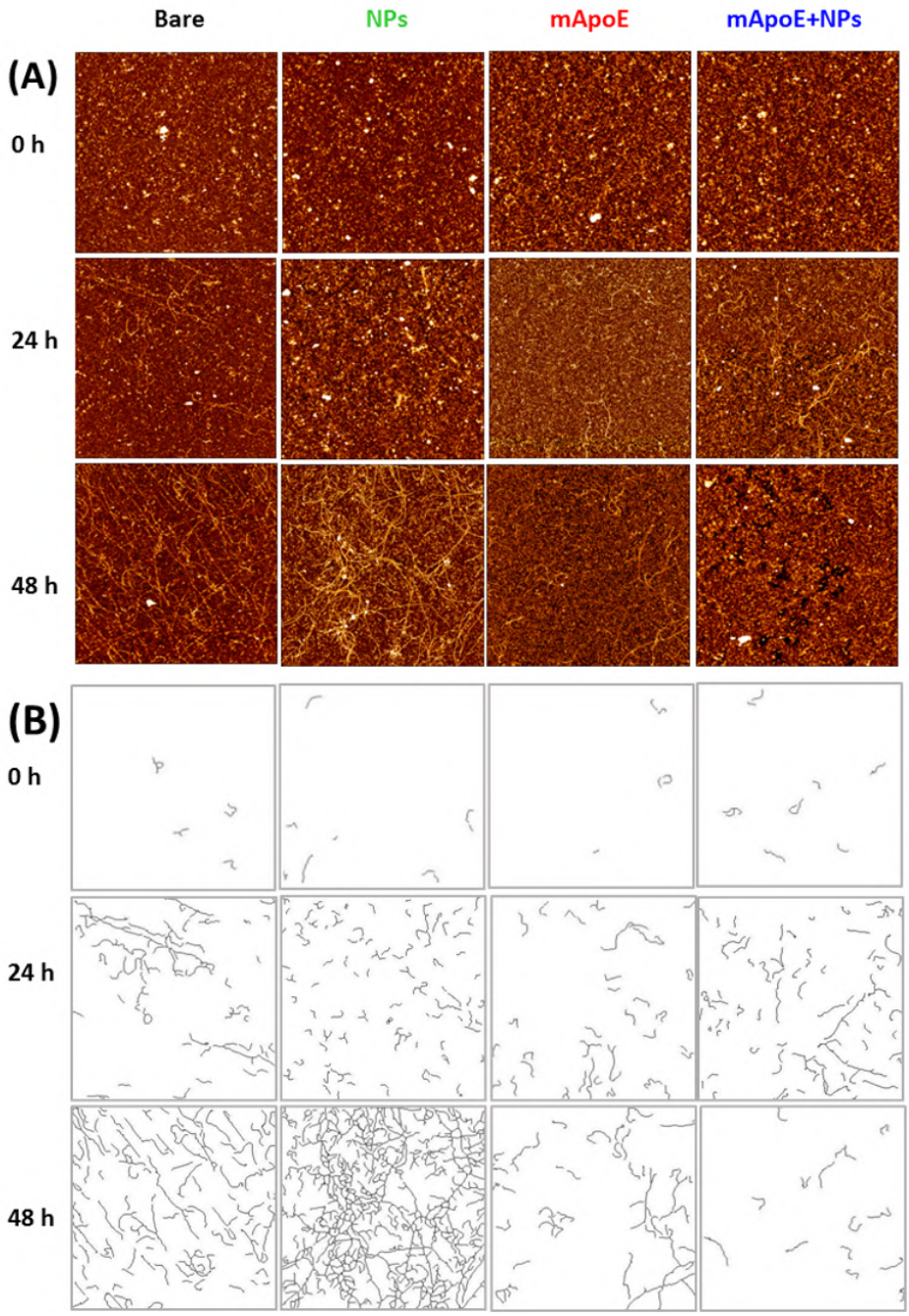

Figure A.7.: **AFM images relative to the effect of non functionalized AuNP, free mApoE, and mApoE functionalized AuNP on the fibrillar aggregation of** $A\beta_{1-42}$. A) Representative AFM images (4x4 μm^2, 1024 x 1024 pixel, Z-scale 10 nm) of typical aggregation patterns of bare $A\beta_{1-42}$ amyloid fibrils (first column), of $A\beta_{1-42}$ amyloid fibrils in presence of non functionalized AuNP (second column), or free mApoE (third column), or mApoE functionalized AuNP (fourth column). Data taken at three different incubation times: t=0 h (first line), t=24 h (second line) and t=48 h (third line). B) Post-processed analysis of the AFM images reported in A. The lines represent the resulting deposited fibrils as selected by the indicated software procedure.

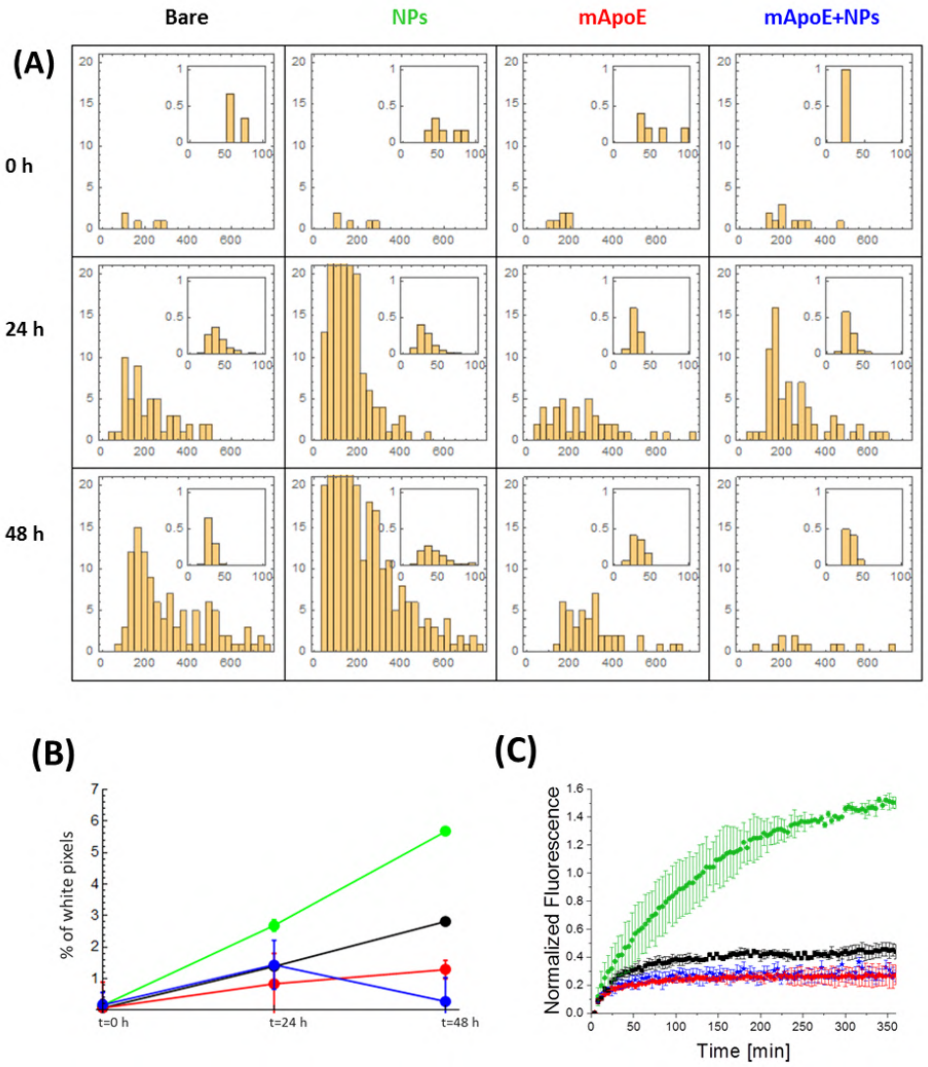

Figure A.8.: **Quantitative analysis of the effect of non functionalized NP, mApoE, mApoE functionalized NP on the aggregation of** $A\beta_{1-42}$. A) Statistical distribution of fibrils length (main figures) and width (insets) as obtained from the analysis of AFM data reported in Fig.A.7B. The values in the histograms are reported in nm. B) Quantitative analysis of AFM images reported in A. The percentage of deposited fibrils is plotted as a function of time of aggregation. Data obtained for $A\beta_{1-42}$ bare in solution (black dots) or in the presence of non functionalized NP (green dots), free mApoE (red dots), or mApoE functionalized NP (blue dots). C) FITC fluorescence representing the aggregation of 30 nM FITC-$A\beta_{1-42}$ as a function of the aggregation time: free $A\beta$ (black squares) is compared to $A\beta$ added with non functionalized NP (green circles), free mApoE (red diamonds), or mApoE functionalized NP (blue stars). The normalized fluorescence values are the average over the three experiment repetitions, with error bars corresponding to the pertaining standard errors.

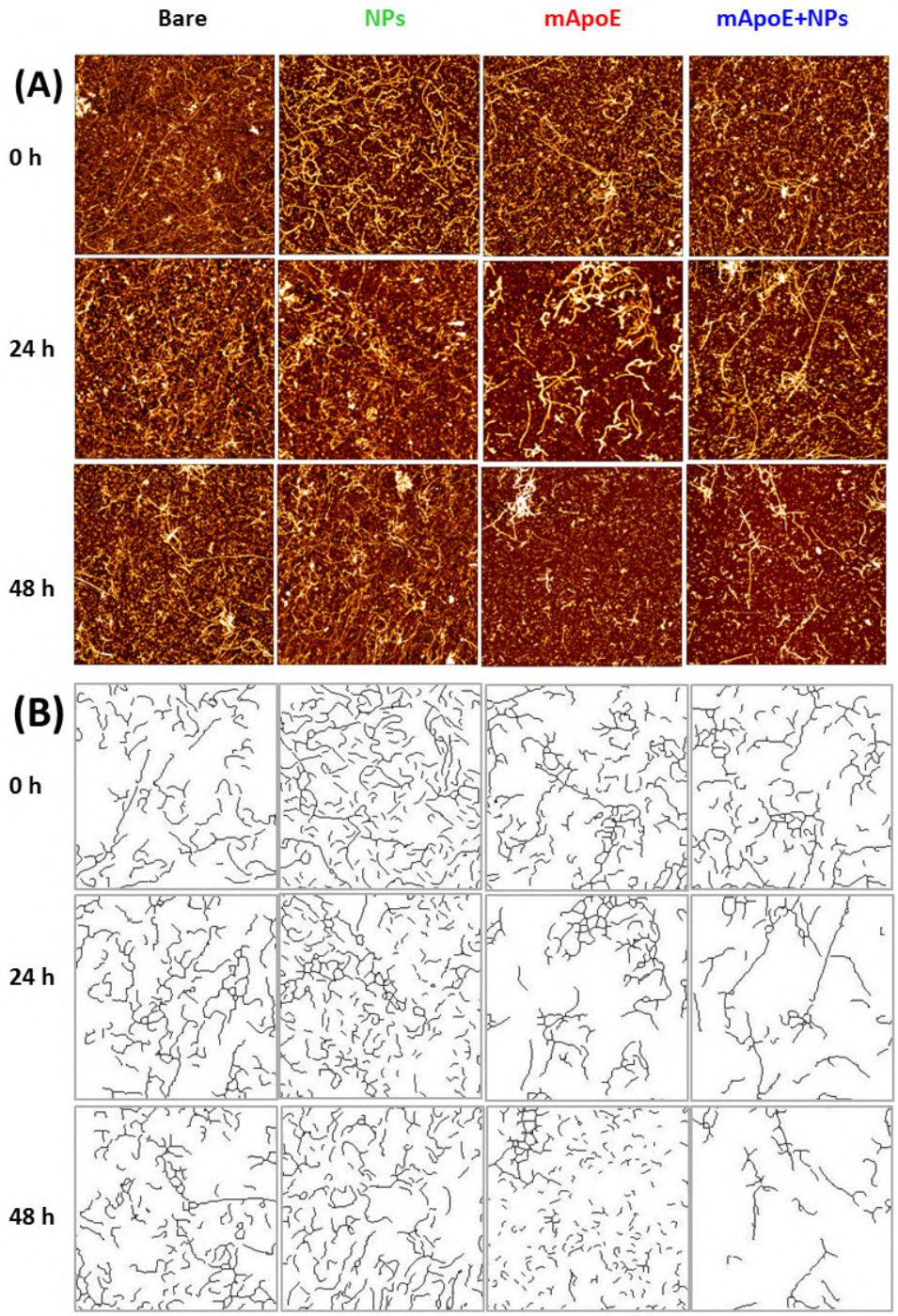

Figure A.9.: **AFM images of the effect of non functionalized AuNP, free mApoE, and mApoE functionalized AuNP on the disaggregation of preformed Aβ fibrils.** (A) Representative AFM images (4x4 μm^2, 1024 x 1024 pixel, Z-scale 10 nm) of typical disaggregation patterns of preformed $A\beta_{1-42}$ amyloid fibrils (first column), of $A\beta_{1-42}$ amyloid fibrils in presence of non functionalized AuNP (second column), or free mApoE (third column), or mApoE functionalized AuNP (fourth column). Data taken at three different incubation times: t=0 h (first line), t=24 h (second line) and t=48 h (third line). B) Post-processed analysis of the AFM images reported in A. The lines represent the resulting deposited fibrils as selected by the indicated software procedure.

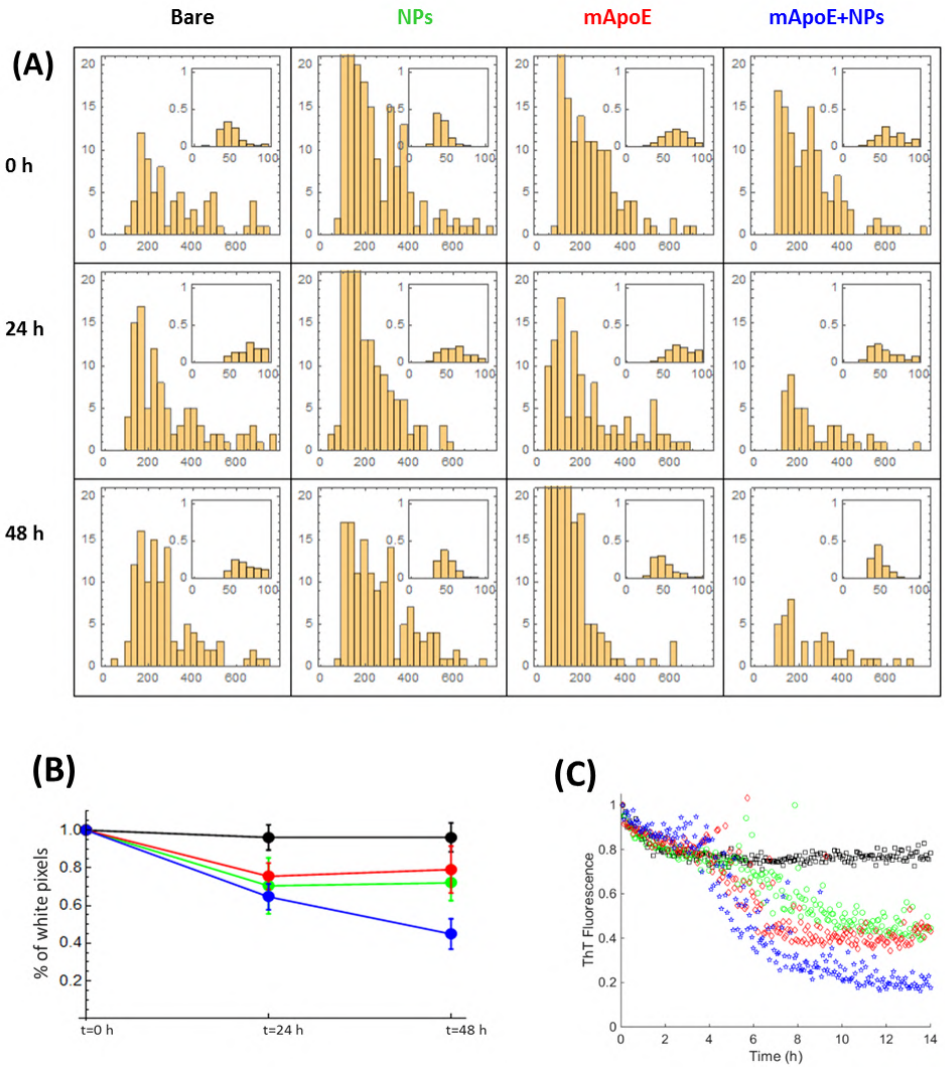

Figure A.10.: **Quantitative analysis of the effect of non functionalized AuNP, free mApoE, and mApoE functionalized AuNP on the disaggregation of preformed Aβ fibrils**. A) Statistical analysis of fibrils length (main figures) and width (insets) as obtained from the AFM data reported in Fig.A.9B. The values in the histograms are reported in nm. B) Quantitative analysis of AFM images reported in A. Percentage of deposited fibrils plotted as a function of time of disaggregation. Data obtained for Aβ bare in solution (black dots) and in the presence of bare NP (green dots), free mApoE (red dots) or AuNPs coated with mApoE (blue dots). The reported percentage of deposited fibrils is normalized with respect the initial value (t=0 h). C) ThT assay-based evaluation of the evolution of preformed fibrils. THT fluorescence plotted as a function of time in 2 μM concentrated pre-fibrillated samples of Aβ stained with 10 mM THT: free Aβ (black squares) is compared to Aβ added with non functionalized NP (green circles), free mApoE (red diamonds), or mApoE functionalized NP (blue stars). The intensities were normalized to the initial (t=0 h) intensity.

A.2 DNA PROTEIN-MEDIATED COLLAPSING AND CROWDING AGENT EFFECT (HNS AND PEG)

It has been suggested that the combination of bridging from the Histone-like Nucleoid Structuring protein (H-NS) and molecular crowding from the cellular environment may play a major role in the physical structuring of the *E. coli* genome [265, 266, 267]. The macromolecules of the bacterial cell occupy a large amount of the total cytosol volume and crowded environments have long been known to compact and stabilize DNA [268, 269]. To study the interplay of this physical factor and the H-NS bridging activity in a clean controlled setting, we analyzed the structural compaction and folding kinetics on short DNA molecules in the presence of both H-NS and crowding induced by poly-ethylene glycol (PEG). We used two different single molecule nanomechanical techniques, MT (Fig.A.11) and AFM. All the data and figures reported here are taken or inspired from [9].

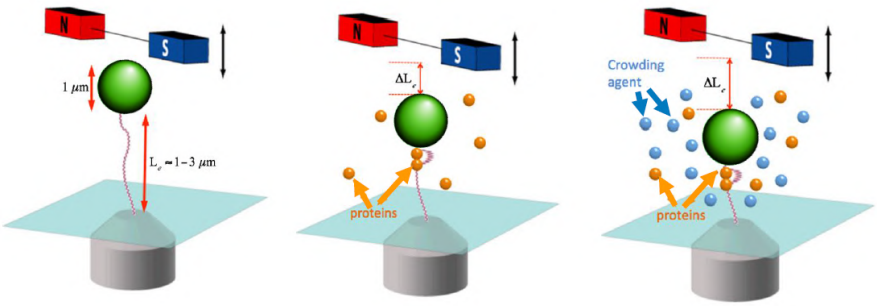

Figure A.11.: Sketch (not in scale) of the MT experiments. One end of a DNA filament is attached to a fixed substrate, while the other end is connected to a magnetic bead. As indicated in the sketches, at a fixed force, the presence of proteins and/or crowding agents reduce the DNA extension by a value of ΔL_e.

H-NS alone showed a step-wise DNA collapse due to the formation of multiple bridges, with a critical unfolding force that was independent from H-NS concentration. Conversely, PEG alone showed an highly concentration-dependent collapse force. We interpreted the two limit cases using the existing theories of loop formation in a pulled chain and pulling of an equilibrium globule. Finally combination of the two agents showed a clear cooperative effect between the activity of the H-NS protein and the depletion force created by PEG. This behaviour can be described by considering both the previously proposed enhancement of the depletion attraction by H-NS coating and the persistence of (modified) loop-forming properties of H-NS. We hypothesized that this double role of H-NS in enhancing compaction while forming specific loops could be important *in vivo* for defining specific mesoscale domains in chromosomal regions. Here, aiming to get a cleaner view on the joint role played by H-NS

and crowders on DNA collapse, we used a single-molecule MT setup to investigate the DNA extension under the effect of externally applied forces, and under an interplay of H-NS (in the regime of the bridging mode) and crowding, induced by the 1500 molecular weight PEG, (PEG1500) at various concentrations. We combined these measurements with AFM imaging techniques, aiming to complement our view of the molecular interaction of DNA with H-NS or PEG.

A.2.1 *SMFS of H-NS and PEG mediated DNA collapsing*

Fig.A.12 reports representative experimental MT force-extension curves showing the effects of various H-NS and PEG concentrations on the DNA collapse (red lines represent the trace of bare DNA, whereas the lines in different blueish or greenish colors illustrate the collapse of DNA molecules at various concentrations of H-NS or PEG respectively).

As shown in Fig.A.12, all the DNA folding curves in the presence of H-NS overlapped approximately at the same critical force ($F_{bridging} = 0.15 \pm 0.01$ pN, reported in the inset as blue circles), indicating that the bridging phenomenon was not affected by protein concentration.

Conversely, the DNA condensation induced by PEG, strongly depended on the concentration of the crowding agent, as shown in Fig.A.12. The force of DNA collapse in the presence of PEG was increasingly shifted toward higher values for higher PEG concentrations (reported in the insets of Fig.A.12).

The different behaviour of condensation in the presence of PEG and H-NS was also witnessed by temporal evolution of the DNA extension at a fixed force. Fig.A.13 shows representative temporal traces of the DNA collapse induced by 30 nM H-NS or 19 % PEG. Overall, in the presence of H-NS we observed a clear multistep feature typical of the formation of multiple loops, while in the presence of PEG we noted an abrupt and quick jump of DNA with no intermediate steps.

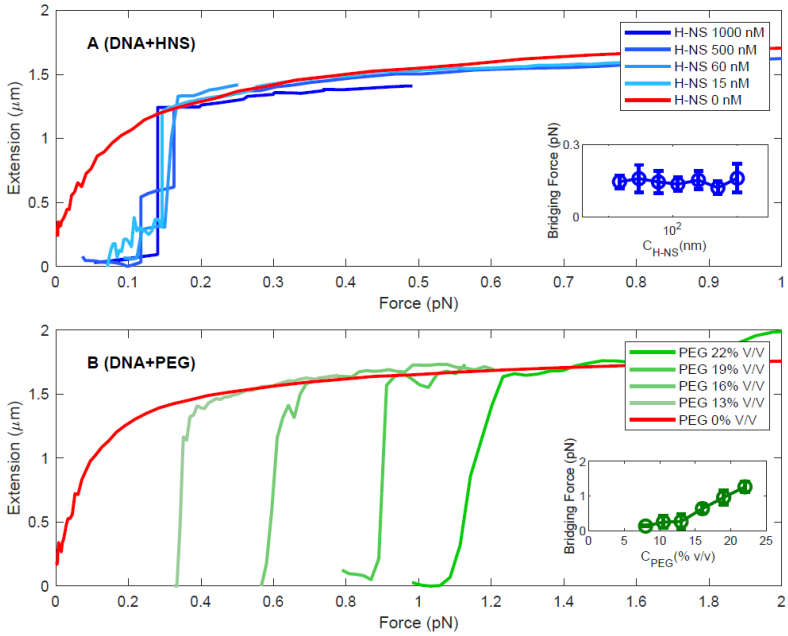

Figure A.12.: The concentration dependence of H-NS bridging-binding mechanism (Top) and the DNA collapse induced by PEG (Bottom) is qualitatively different. Force-extension curves acquired on bare DNA (red lines) and on DNA in the presence of H-NS (blueish lines) or PEG (greenish lines) respectively, at concentrations ranging between 15 nM and 1 µM for H-NS or between 13 % and 22 % v/v for PEG. Insets: bridging force measured as a function of the H-NS or PEG concentrations.

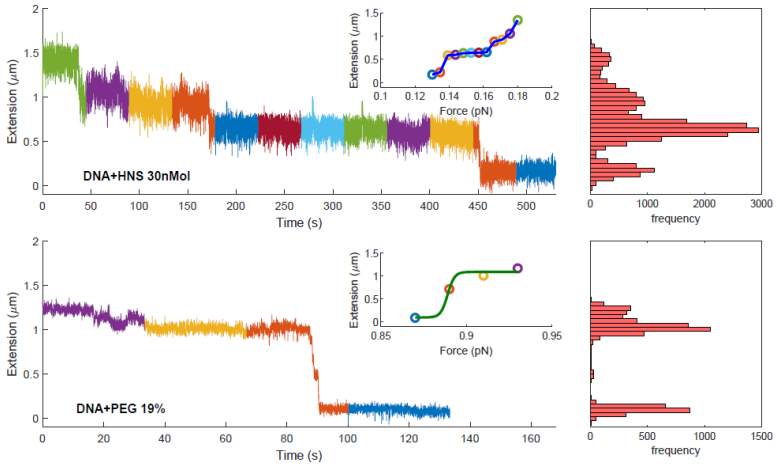

Figure A.13.: Representative temporal traces of the DNA end-to-end extension measured during a DNA compaction event induced by H-NS or PEG: data taken in the presence of 30 nM H-NS (Top, forces decreasing from 0.18 pN to 0.13 pN) and PEG 19 % v/v (Bottom, forces decreasing from 0.93 pN to 0.87 pN). Right panels: histograms on the end-to-end extension of DNA reported in the temporal traces of the left panels.

A.2.2 *AFM imaging shows H-NS and PEG interactions with the DNA chain*

To qualitatively visualize the characteristics of the DNA collapse in presence of H-NS or PEG, we collected and analyzed several high resolution AFM images. Some representative example are shown in Fig.A.14 for H-NS:DNA and in Fig.A.15 and Fig.A.16 for PEG:DNA complexes respectively.

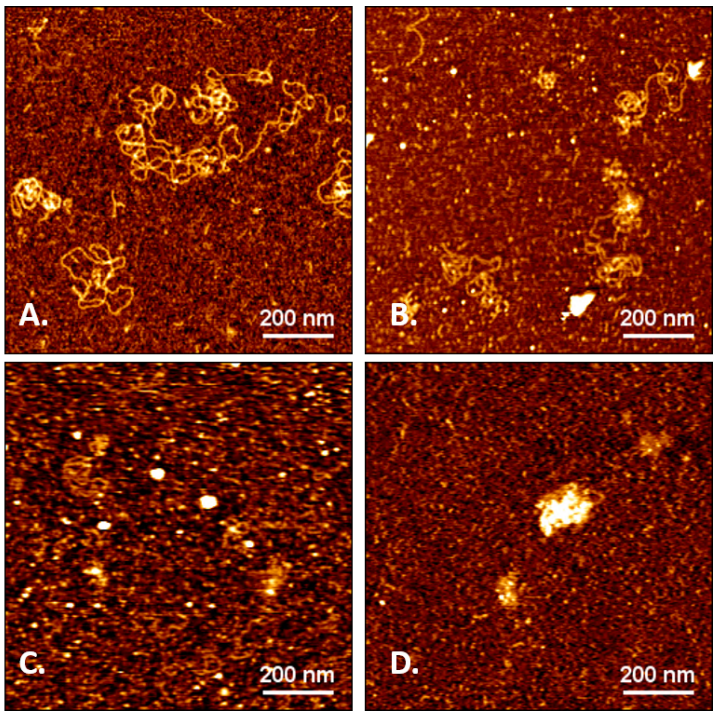

Figure A.14.: AFM images of HNS-induced DNA collapse in the presence of 0 (A), 50 nM (B), 100 nM (C) and 500 nM (D) H-NS respectively, at the DNA concentration of 0.3 ng/ µl. (images acquired in H-NS bridging buffer 10 mM Tris-HCl at pH 7.5, 60 mM KCl, 10 mM $MgCl_2$, 1x1 μm^2, 256 x 256 pixel, Z-scale 1.5 nm).

As illustrated in Fig.A.14, Fig.A.15 and Fig.A.16, increasing the H-NS or PEG concentrations, we observe a clear progressive condensation of DNA.

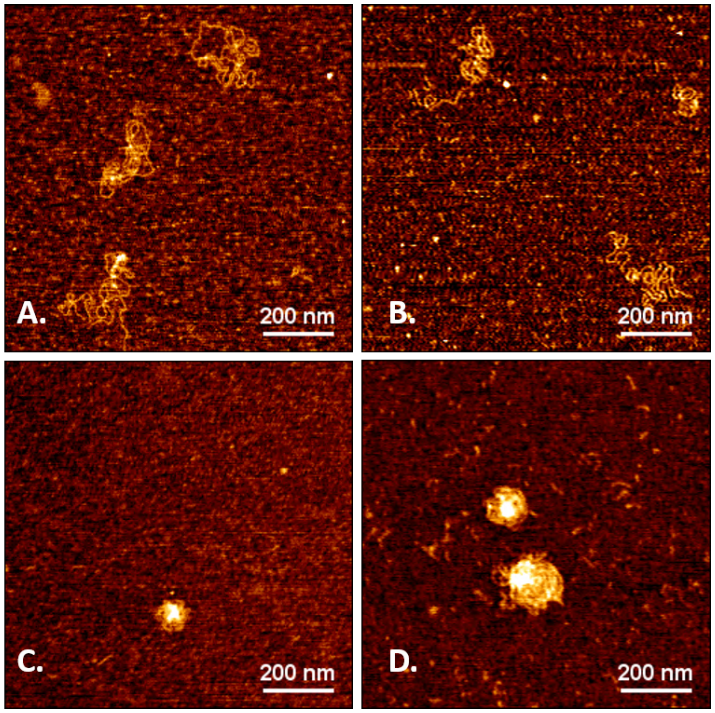

Figure A.15.: AFM images of PEG-induced DNA collapse in the presence of 0 (A), 7.5% *v*/*v* (B), 10% *v*/*v* (C) and 15% *v*/*v* (D) PEG respectively, at the DNA concentration of 0.3 ng/µl. (Images acquired in the same conditions reported in Fig.A.14).

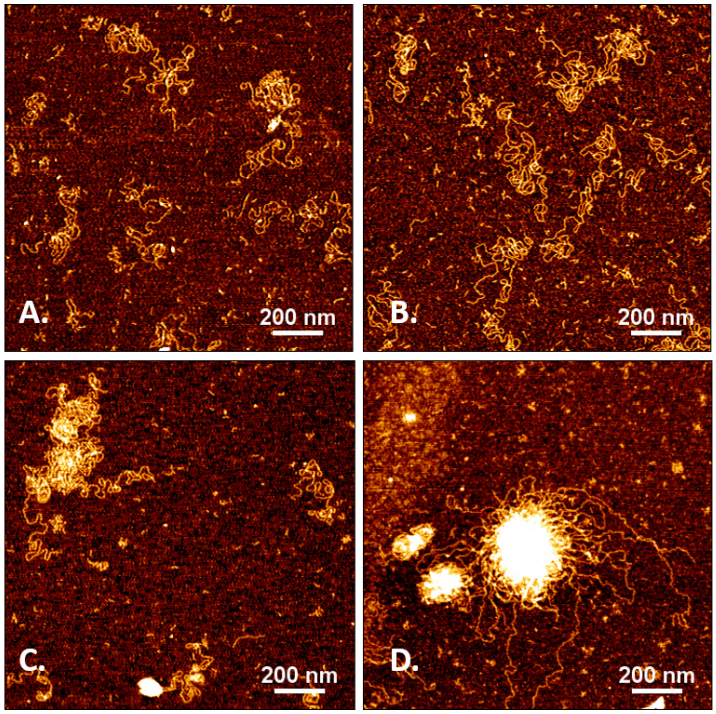

Figure A.16.: AFM images of PEG-induced DNA collapse in the presence of 0 (A), 7.5% *v*/*v* (B), 10% *v*/*v* (C) and 15% *v*/*v* (D) PEG respectively, at the DNA concentration of 1.2 ng/µl. (Images acquired in the same conditions reported in Fig.A.14).

A.2.3 *Cooperative effects of H-NS and PEG*

After the investigation of the individual behaviour of the DNA extension in the presence of H-NS or PEG, we focused on the combined effects of the protein and the crowding agent on DNA condensation. These experiments, simultaneously exploring the nanomechanics of the interaction between DNA and H-NS plus PEG, simulated more realistic conditions, closer to the ones typical of a cellular environment.

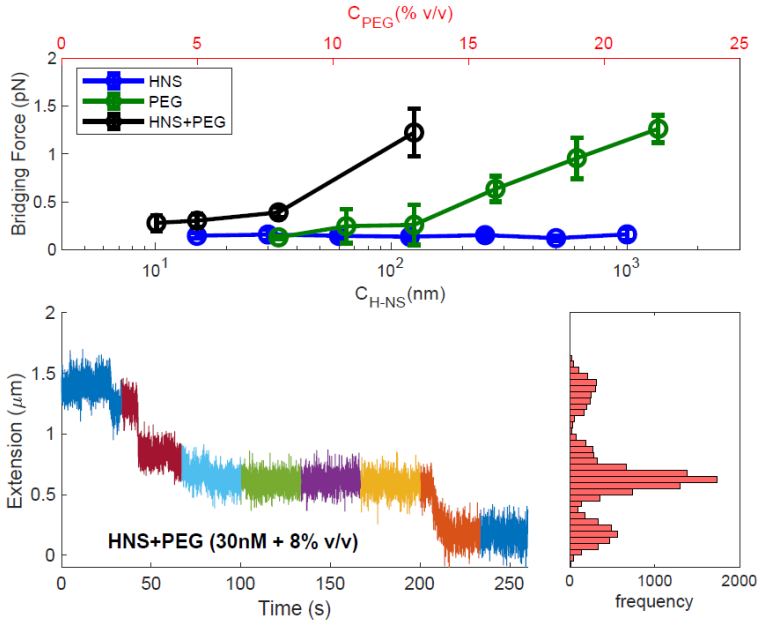

Figure A.17.: Cooperative effects on the DNA bridging and collapse folding force in presence of H-NS and PEG. (Top) Bridging force F_{bridg} as a function of H-NS concentration (blue circle, lower horizontal axis), and folding force F_{fold} as a function of PEG concentration (green circle, upper horizontal axist). Data taken in presence of both H-NS and PEG (black circles) at the same H-NS and PEG concentrations indicated on the upper and lower horizontal axis. Error bars represent the standard deviations. (Bottom) Representative temporal trace of a force-extension experiment during DNA folding in presence of 30 nM H-NS and 8% v/v PEG. Right panel: corresponding statistical distribution of the DNA extension.

We used the qualitatively different dependency of the critical forces on PEG volume fraction reported in the inset of Fig.A.12 for H-NS or PEG alone, as a test for cooperative effects in the presence of both H-NS and PEG (Fig.A.17). The plot clearly indicate that the value of the critical force (F_{bridg}) to unwind a PEG-triggered globule increase drastically in the presence of also H-NS. This is in line with the hypothesis formulated by previous studies, that H-NS affected the depletion interaction induced by PEG, helping DNA collapse [270].

Additionally, from the temporal traces recorded during the DNA condensation we could verify directly whether the steps triggered by loop formation still occurred in the presence of both H-NS and PEG. Indeed, we report in Fig.A.17 a representative temporal trace and its corresponding statistical distribution of the DNA extension in the presence of both H-NS and PEG. Here, the kinetics of DNA folding induced by H-NS and PEG showed very similar feature to the kinetics of DNA bridging mechanism of H-NS alone (see Fig.A.13, Top). The presence of a multistep temporal trace was a clear signature of the protein binding, which is different from the crowding effect alone (see Fig.A.13, Bottom). Hence, we conclude that H-NS appeared to play the double role of enhancing depletion forces induced by PEG, and specifically forming loops. This role could be important in an *in vivo* context, to help maintaining separate domains, which are *programmed* on the chromosome by the specific binding action of H-NS, even in the presence of strong depletion forces collapsing the chromosome. Thanks to the *pinning* effects of H-NS maintaining specific loops, a collapsed *in vivo* chromosome could maintain compartmentalization and achieve very different properties from an equilibrium globule.

The creation of an artificial extracellular matrix (ECM), opportunely tunable for different type of cells is one of the most important achievement for the regenerative medicine as well as for in-vitro cellular studies. The cell microenvironment plays a pivotal role in mediating cell adhesion, survival, and proliferation in physiological and pathological states. The glycosylation of ECM proteins remains, however, largely unexplored. The charachterization of a chemoselective glycosylation approach for collagen matrices was performed by means of AFM imaging.

AFM for glyco-functionalized matrix imaging

Atomic Force Microscopy (AFM) measurements were performed in air using a Multimode 8 AFM (Bruker Corporation, Santa Barbara, CA, USA). Images were acquired in peak force tapping mode (PeakForce-Quantitative Nano-Mechanics, PF-QNM). Collagen films were immobilized on a glass surface by letting evaporate a drop of water. V-shaped Scan Asyst Fluid + (0.7 N/m) cantilevers were used. Data processing was performed using the commercial Nanoscope Analysis software (Bruker Corporation, Santa Barbara, CA, USA).

AFM analysis (Fig.A.18) reveals that both neoglycosylated films (Samples 3 and 4) have well-defined fibrillary structures, and on the contrary, the untreated control shows amorphous structures. The fibrillary organization of collagen is a fundamental feature in cell culture and tissue engineering.

As highlighted in Fig.A.18 and A.20, Sample 3 and Sample 4 present well-defined fibrils, having respectively pitches of 68 ± 5 nm (Sample 3) and 72 ± 2 nm (Sample 4), regardless of the width of the fibril.

The roughness $R_q$of the samples was calculated in according with the standard formula

$$R_q = \sqrt{\frac{\sum z_i^2}{N}}$$

by using the available tool of the commercial Nanoscope Analysis software (Bruker Corporation, Santa Barbara, CA). The images were splitted into 25 different regions and the resulting surface roughness are evaluated and averaged. Despite the presence of the fibrils, the roughness of the CT sample ($R_q = 196 \pm 44$ nm) is approximately two times bigger than the neoglycosylated collagen films ($R_q = 74 \pm 15$ nm for Sample 3 and $R_q = 111 \pm 25$ nm for Sample 4). Also the R_a for each sample was evaluated by the means of the standard formula

$$R_a = \frac{1}{N} \sum_{j=1}^{N} |Z_j|$$

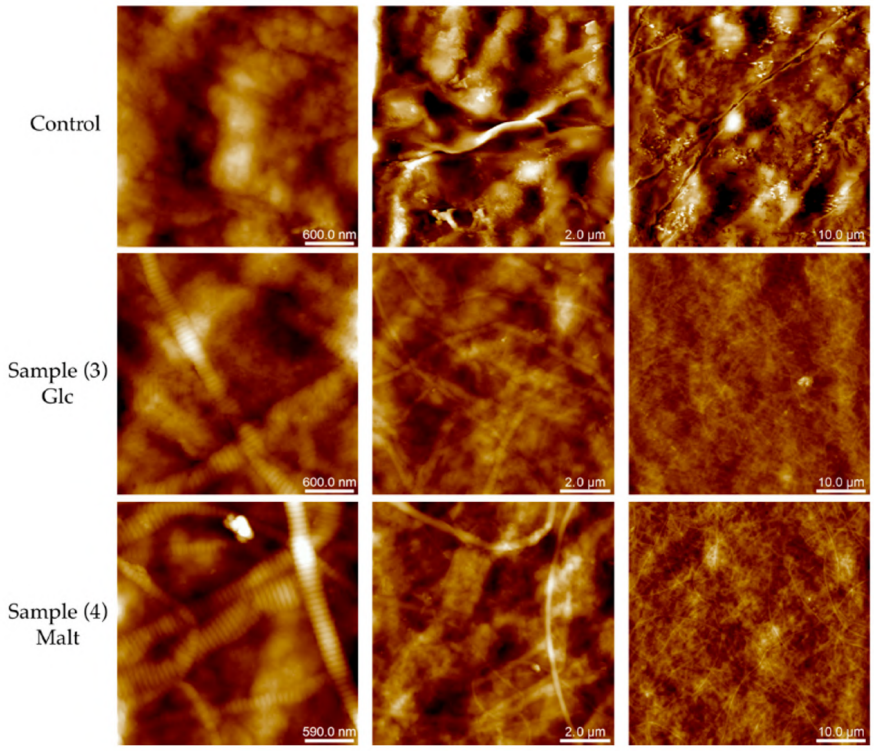

Figure A.18.: Representative AFM topography image of collagen films acquired in PeakForce mode in air. Different scan sized images of collagen matrix without further modification, glucose (line 2) and maltose (line 3) neoglycosylated collagen films were collected. From left to right: column 1 (3 x 3 μm^2, 512 z 512 pixel, Z-scale 200 nm), column 2 (10 x 10 μm^2, 512 x 512 pixel, Z-scale 500 nm), and column 3 (50 x 50 μm^2, 512 x 512 pixel, Z-scale 1.4 µm). Fig taken from [3].

reporting values showing the same behavior than the ones found for the R_q (R_a = 165 $\pm$ 36 nm for CT, R_a = 59 $\pm$ 12 nm for Sample 3 and R_a = 88 $\pm$ 18 nm for Sample 4). All the values of roughness here reported are mean value $\pm$ st. dev.

Collagen molecules are packed in a quarter-staggered fashion which gives rise to a repeating banding pattern, the so-called D-periodicity or D-band, of about 67 nm [271, 272]. It has been reported in the literature that the transverse D-banding periodic pattern is a key player with respect to fibril mechanical properties, which significantly impacts cell-collagen interactions and is correlated with pathological conditions. In nature, the structural, mechanical, and functional features of native collagen in the ECM are strongly linked to the contribution of several factors. These include a) the interactions with other ECM proteins and soluble proteins (interactors), b) the interaction with cell surface receptors, and c) post-translational modifications. Data reported in the literature on collagen triple helix formation [273] suggests that the glycosylation of collagen chains has an

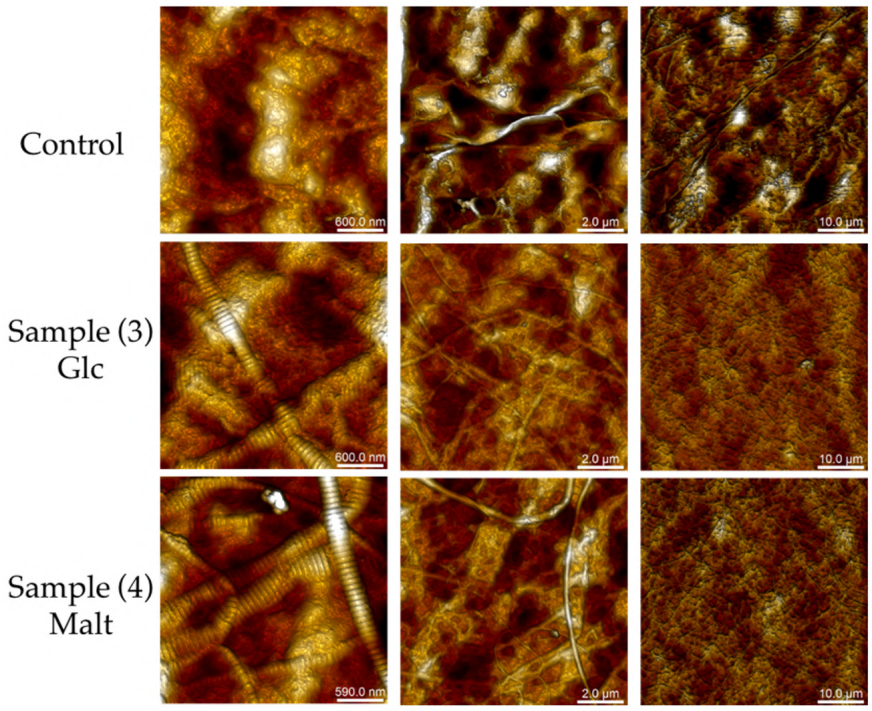

Figure A.19.: 3D Lighting (Nanoscope Analysis software -Bruker Corporation, Santa Barbara, CA) image of collagen matrices acquired in PeakForce mode in air reported in the Figure A.18. Different scan sized images of collagen matrix without further modification (line 1), glucose (line 2) and maltose (line 3) neoglycosylated collagen matrices were collected. From left to right: column 1 (3 x 3 μm^2, 512 x 512 pixel, Z-scale 200 nm), column 2 (10 x 10 μm^2, 512 x 512 pixel, Z-scale 500 nm), and column 3 (50 x 50 μm^2, 512 x 512 pixel, Z-scale 1.4 µm). Fig taken from [3].

important contribution to surface roughness variations in cell–ECM and ECM–ECM interaction [274, 275]. As verified by AFM analysis, morphological changes in neoglycosylated collagen could have an influence on both the inter-molecular and inter-fibrillar interactions of the triple-helical domain of collagen films, contributing also to the improved biological activity of the produced films.

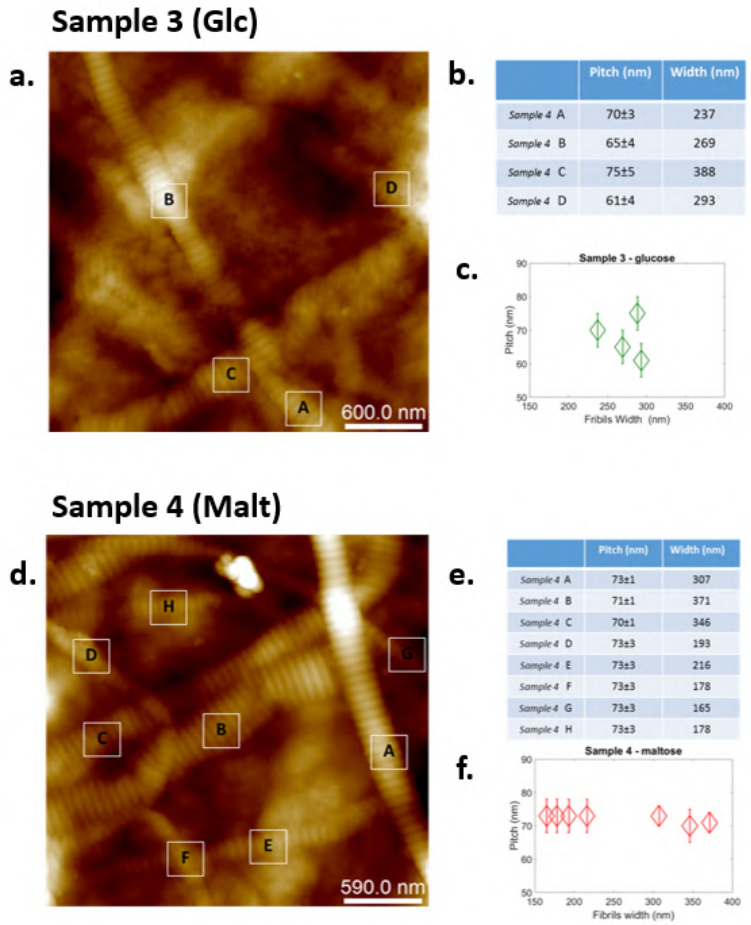

Figure A.20.: 3 x 3 μm^2 AFM topography images of Sample 3 - Glc (a) and Sample 4 - Malt (d) reported in Fig.A.18 with the indication of the chosen fibrils for the pitch analysis, (b-e) the pitch and the width of each fibrils and (c-f) the absence of correlation of the pitch with the fibril's width. Figure taken from [3].

B

AFM-BASED ELASTICITY MEASUREMENTS

B.1 COLLAGEN MATRIX

B.1.1 *Endogenous LOX Secreted by ccRCC Cells Modifies the Collagen Matrix Stiffness*

Human clear cell renal cell carcinoma (ccRCC) tumour progression has been correlated with Lysyl Oxidase (LOX) capacity to crosslink collagen and increase extracellular matrix (ECM) stiffness. The mechanical effects of secreted LOX on collagen-coated gels was quantified using the atomic force microscopy (AFM) and calculating the values of the Young modulus (see Sec.1.3.3). The ccRCC conditioned media, in which LOX is present, induced a significant increase of collagen matrix stiffness evidenced by the Young modulus that is higher by a factor of 1.30 ± 0.32 (means $\pm$ SD) with respect to the value of collagen matrices treated with fresh control media (Fig. B.1). The presence of bAPN LOX inhibitor in ccRCC-conditioned media prevented collagen matrix from becoming stiffer, resulting in a non significant variation of Young modulus (0.90 ± 0.23) with respect to the control media treatment. These data were confirmed even by using the conditioned medium of LOX silenced ccRCC cells (Fig. B.1). Overall, the stiffness measurements highlighted that the active LOX constitutively produced by ccRCC increased the collagen matrix stiffness.

Collagen Matrix Stiffness Measurements by AFM

Collagen matrices were prepared as described in the following paragraph. Briefly, 15µL of 7.5% acrylamide and 0.4% bisacrylamide (Sigma-Aldrich) gel solution was delivered on round 25-mm glass coverslips, and after polymerization, 300 mL of Sulfo-SANPAH (Sigma-Aldrich) was added to gel surface. After 2 hours, 1 mL of 1 mg/mL of fresh collagen solution was added to the gel surface and incubated overnight at 37°C. Phosphate-buffered saline (PBS) rinsed collagen-coated matrices were then treated overnight at 37°C with (i) control medium, ii) control medium of ccRCC primary cell culture plus 500 mM bAPN LOX Inhibitor (Sigma-Aldrich), iii) conditioned medium of ccRCC primary cell culture, iv) conditioned medium of

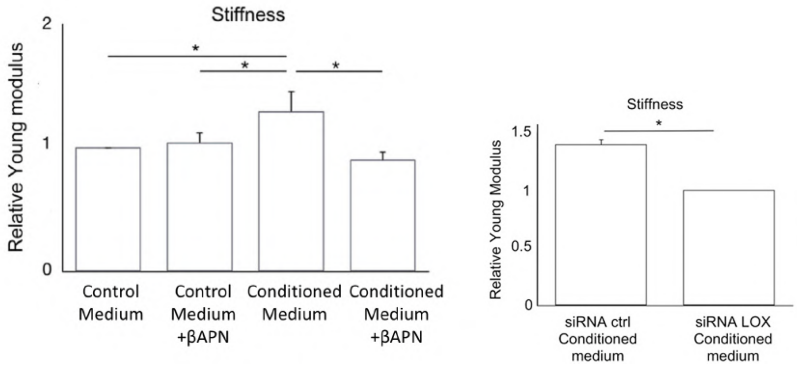

Figure B.1.: (Left) Evaluation of collagen matrix stiffness after treatment with control medium, control medium containing LOX inhibitor βAPN, conditioned medium of ccRCC primary cultures expressing HIF-1a, and the same ccRCC conditioned medium containing bAPN. Data were represented as fold-change with respect to control medium considered equal to 1. *P < 0.05 (t-test). (Right) Evaluation of collagen matrix stiffness after treatment with conditioned medium of LOX silenced and control ccRCC primary cultures. Data are expressed as fold-change with respect to siRNA LOX conditioned medium considered equal to 1. Fig taken from [6].

ccRCC primary cultures plus 500 mM bAPN, v) conditioned medium of ccRCC cells treated with control siRNA, and vi) conditioned medium of ccRCC cells treated with LOX siRNA for Fig.B.1, Right). Stiffness measurements of differently treated collagen matrixes were performed in PBS solution using standard AFM Nanowizard II (JPK, Berlin, Germany). The force-displacement curves between contact point and the 2000-nm deformation depth have been analysed and fitted by means of Hertz Model. Young modulus was represented as fold-change with respect to control considered equal to 1.

B.1.2 *The 1ALCTL and 1BLCTL isoforms of Arg/Abl2 induce fibroblast activation and extra cellular matrix remodelling differently*

The activated fibroblasts (myofibroblasts), which are abundant in the fibrotic tissue closed to some kind of cancer cell, play a role in the creation of a cancer cell favourable environment, characterised by abundant extracellular matrix (ECM) secretion [5]. The myofibroblasts remodel this tissue through secreted molecules and modulation of their cytoskeleton and specialized contractile structures. The non-receptor protein tyrosine kinase Arg has the unique ability to bind directly to the actin cytoskeleton, transducing diverse extracellular signals into cytoskeletal rearrangements. Here, we analysed the 1ALCTL and 1BLCTL Arg isoforms in Arg−/− murine embryonal fibroblasts (MEF) cell line, focusing on their capacity to activate fibroblasts and to

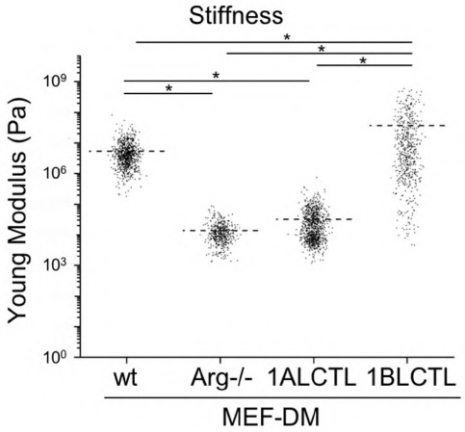

Figure B.2.: Evaluation of stiffness of the MEF-DM. Dot plot shows the stiffness expressed as Young modulus (Pa) of the different matrices, each dot represents a measure performed in different positions of MEF-DM during independent experiments (n=3). Fig taken from [5].

remodel ECM, including a stifness different modulazion. The results obtained showed that Arg isoform 1BLCTL has a major role in proliferation, migration/invasion of MEF and in inducing an environment able to modulate tumour cell morphology, while 1ALCTL isoform has a role in MEF adhesion maintaining active focal adhesions. On the whole, the presence of Arg in MEF supports the proliferation, activation, adhesion, ECM contraction and stiffness, while the absence of Arg affected these myofibroblast features.

As fibronectin and collagen I deposition is relevant in determining matrix stiffness, we assessed the relation between the produced ECM and their respective stiffness after removing MEF cells. Fig. B.2 showed a decreased stiffness of ECM produced by Arg$-/-$ MEF with respect to wt MEF. The 1ALCTL Arg isoform was unable to elicit the production of matrix with stiffness at the level of wt MEF, while 1BLCTL transfection elicited the production of ECM with the highest stiffness above all the samples.

The ECM stiffness is due to the amount of collagen and fibronectin fibres, their cross-link and ECM morphology. The specific role of Arg in producing fibronectin and collagen matrix turned out to be also significant in producing a specific matrix framework able to modulate the tumour cell morphology.

In the last two decades, several efforts on single cell force spectroscopy (SCFS) aimed to quantitatively measure the mechanical properties of living cells and relate them to cell structure and function. It has been demonstrated that the cells mechanical properties are closely connected to several important cellular functions such as differentiation, adhesion, motility, proliferation, internal molecular transport, or signal transmission [276, 277]. Among other techniques, the AFM-based SCFS has proved to be a valuable tool for the quantitative characterization of static and frequency-dependent mechanical properties of micro-structures, such as cells, thanks to its ability to sense and apply nanoscale forces in near physiological environments [26, 28].

Several studies reported the correlation between cells elasticity and their patho-physiological state, including cancer diseases [28, 29, 278, 279, 280]. Usually, low rigidity of cancer cells was suggested as an indicator for cancer diagnosis [28, 29]. However, there are results demonstrating no change [281] or even increase of rigidity [282] with malignant carcinoma development.

Here, three different studies of cellular elasticity based on AFM-based SCFS technique are reported.

Materials and Methods: SCFS on cell lines

Prior to experiments, cells were cultured on Poly-D-lysine functionalized glass (for ETNK cell lines), on Poly-L-ornithine 1:10 (for MEC cell lines) or directly on non-functionalized standard petri dishes in their culture medium. Cells were then incubated in the CO_2 incubator for at least 2 hours, then washed very carefully 2x times with sterile PBS solution (pH 7.4).
All SCFS measurements were carried out using standard AFM Nanowizard II (JPK) working in force spectroscopy mode. To prevent significant changes in morphology or biochemistry of living cells outside the incubator, each sample was measured within 2 hours. During that time, at least two force maps (8x8 pixels grid, scan size of 8x8 μm^2) in the center of the cells were recorded. The force set point was set at $0.5 - 1$ nN and the approach and retract speeds were kept at $2\mu m/s$ ($4\mu m/s$ for ETNK cell lines).
For the exemplify measurements reported for pancreatic cells lines in Fig.B.9 and B.10, 20x20 pixels grids $\sim$ 40x40 μm^2 were adopted.
The cantilevers: DNP S10 - D (Nominal Constant 0.06 N/m, Bruker Corporation) or MLCT-BIO-E (Nominal Constant 0.01 N/m, Bruker Corporation) were calibrated before the SCFS measurements both in air and in PBS solution using Thermal Noise method. The force-displacement curves between contact and 500-nm deformation depth have been analysed. Cell elastic properties were evaluated quantita-

tively through the Young's modulus, with the Hertz-Sneddon contact mechanics for a paraboloidal tip.

Materials and Methods: AFM imaging for fixed MEC1 cells

AFM-imaging measurements were performed in air using a Multimode 8 AFM (Bruker Corporation, Santa Barbara, CA, USA). Images were acquired in peak force tapping mode (PeakForce-Quantitative Nano-Mechanics) (50 x 50 μm^2, 256 x 256 pixels). $MEC1^{HS1KO}$ and $MEC1^{HS1UT}$ cells were incubated on collagen films and fixed by using paraformaldehyde 4% and refrigerate overnight at 4°C. Prior to the imaging, specimens were washed two times with MilliQ water and letting evaporate the fluid. V-shaped Scan Asyst Fluid+ (0.7 N/m) cantilevers were used. Data processing was performed using the commercial Nanoscope Analysis software (Bruker Corporation, Santa Barbara, CA, USA).

B.2.1 *ETNK1-mutated cells lines*

ETNK1 kinase is responsible for the phosphorylation of Ethanolamine (Et) to Phosphoethanolamine (P-Et). P-Et plays a critical role in the Kennedy pathway, representing the main metabolic route by which mammalian cells synthetize the two most abundant cell-membrane phospholipids: phosphatidylethanolamine and phosphatidylcholine [7].

A cells-elasticity study has been performed on ETNK1-WT, ETNK1-N244S mutations and ETNK1-KO cells lines, showing a similar rigidity (Fig.B.4) to asses that the ETNK1 mutations not significantly affected cell-membrane composition. The also Phosphoethanolamine (+p-Et in Fig. B.4) phosporilation has been charactherized (only for ETNK1-WT and ETNK1-KO lines, in Fig. B.3).

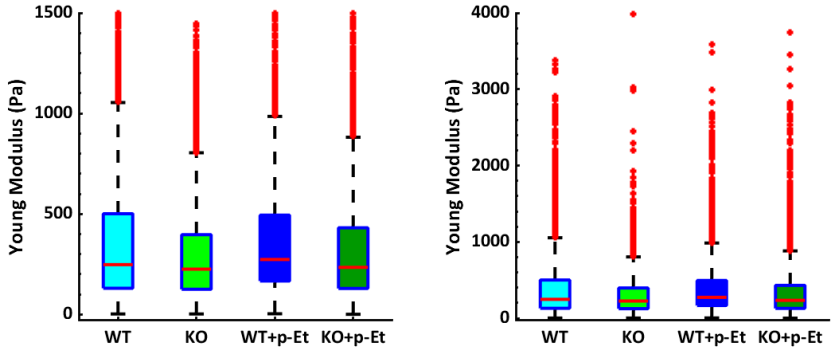

Figure B.3.: Young's modulus of elasticity calculated for ETNK1-WT (light blue), KO (green) lines in the presence of the inhibitor p-ET (dark blue for ETNK1-WT and dark green for ETNK1-KO). Boxes represent the interquartile range; whiskers delimit the 25th and 75th percentile range, red crosses represents the outliers values.

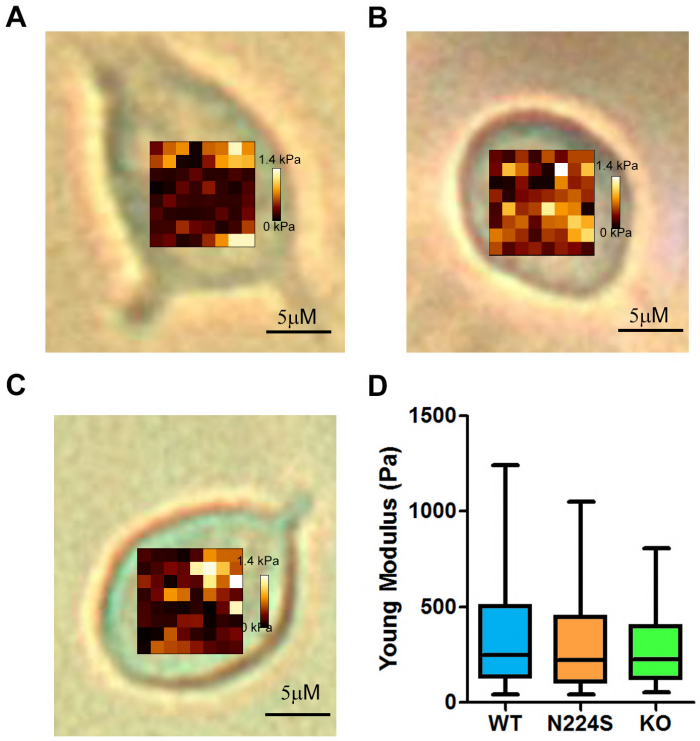

Figure B.4.: A-C) Exemplificative AFM force maps (8x8 pixels grid) of ETNK1-WT (A), ETNK1-N244S (B) and ETNK1-KO (C) cells. Pixel color code is scaled between 0 (black) and 1.4 kiloPascal (white). D) Young's modulus of elasticity calculated for ETNK1-WT (blue), N244S (orange) and KO (green) lines. Boxes represent the interquartile range; whiskers delimit the 5th and 95th percentile range. Figure taken from [7]

B.2.2 *Chronic Lymphocytic Leukaemia Cells*

Chronic lymphocytic leukaemia (CLL) is a B cell lymphoproliferative disorder and represents one of the most common adult leukaemia, with an incidence ranges between 4 to 6 cases per 100,000 people per year in Europe and US. CLL is a very heterogeneous disorder and nowadays there are no prognostic factors able to discriminate CLL patients who will eventually progress at the diagnosis [283].

CLL cells traffic between peripheral blood, bone marrow and secondary lymphatic tissues where interact with the microenvironment [284]. These processes are affected by the mechanical-forces present in the environment and by the capability of the cells to sense the forces. Deciphering the nature and the molecular mechanisms of this complex cross-talk could lead to the identification of new prognostic factors and treatments that operate through novel mechanisms.

It has been demonstrated that Hematopoietic-cell-specific Lyn-substrate-1 (HS1) protein [285] is a cytoskeletal regulator and a prognostic factor in CLL [286], proving that interfering with HS1 expression/phosphorylation impacts on the progression and homing of CLL cells. In particular, previous studies found that patients having an active-HS1 (Y397,Y378) have a good prognosis and a cytoskeletal activity similar to healthy B cells, while those carrying inactive-HS1 show an impaired cytoskeletal activity [287, 288].

The mechanobiology importance has been recognized in different solid tumours: neoplastic cells are capable of sensing the mechanical signals from the microenvironment using mechanosensor receptors, driving cancer progression and invasion. Moreover, CLL cells, like healthy B lymphocytes, are plastic cells constantly re-organizing their cytoskeleton to traffic in and out the tissues and shift from adhesive to non-adhesive phenotype by a sequential engagement of adhesion molecules, chemokine and cytoskeletal proteins. These processes are all affected by the mechanical-forces present in the microenvironment (sheer stress, substrate rigidity). Despite, for healthy B lymphocytes, this plasticity is necessary to perform their immunological functions, the same mechanism might might also drive the inset and progression of CLL.

To further study HS1 role in CLL a knocked-down HS1 mutant in a CLL cell line (MEC1) was produced. By several network analysis on $MEC1^{HS1-KO}$ vs $MEC1^{HS1-UT}$ it was found that HS1-KO significantly affects the expression of molecules involved in: cell motility, adhesion, cell-cell communication, focal adhesion formation [288]. To study the nano-mechanical properties of those cells AFM-based single cell force spectroscopy was performed (Fig.B.5 for $MEC1^{HS1-KO}$ and Fig.B.6 for $MEC1^{HS1-UT}$ cell lines).

By elasticity analysis we found that $MEC1^{HS1KO}$ cells were sensitively less stiff if compared with $MEC1^{HS1-UT}$ cells

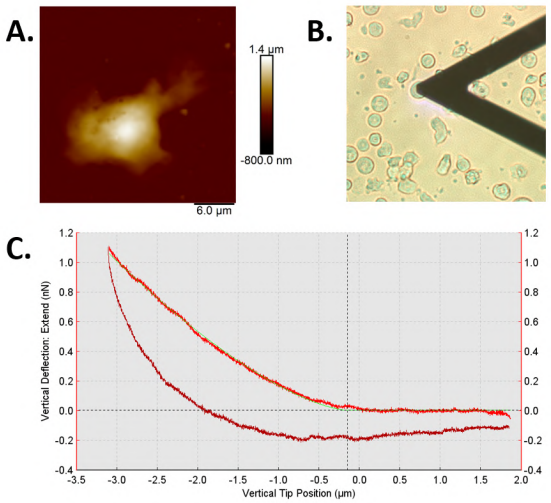

Figure B.5.: A typical experiment of $MEC1^{HS1-KO}$ cell line elasticity. A) AFM image of a fixed KO cell (50 x 50 μm^2, 256 x 256 pixels, z-bar 2.8µm, PeakForce in Air). B) Optical image of cells in the field of view of AFM camera. C) A typical force-vs-extension curve for KO cells (trace in red, retrace in dark red and Hertz-Sneddon model fit for a parabolidal tip in light green).

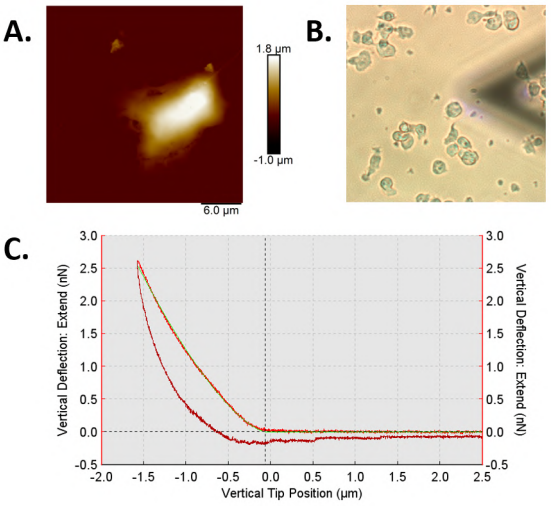

Figure B.6.: A typical experiment of $MEC1^{HS1-UT}$ cell line elasticity. A) AFM image of a fixed WT cell (50 x 50 μm^2, 256 x 256 pixels, z-bar 2.8µm, PeakForce in Air). B) Optical image of cells in the field of view of AFM camera. C) A typical force-vs-extension curve for UT cells (trace in red, retrace in dark red and Hertz-Sneddon model fit for a parabolidal tip in light green).

($E(MEC1^{HS1-KO}) = 1.23 \pm 0.79$ kPa vs $E(MEC1^{HS1-UT}) = 3.40 \pm 2.21$ kPa., Fig.B.7 and B.8). This results demonstrated a putative role for HS1 in regulating the mechanical properties of CLL cells, also in decreasing the cell stiffness and, probably, increasing the cell motility. Moreover, the cell mobility and the minor rigidity allow the cell to

permeate tissues, especially in lymph nodes and bone marrow, where CLL cells start creating the solid tumor.

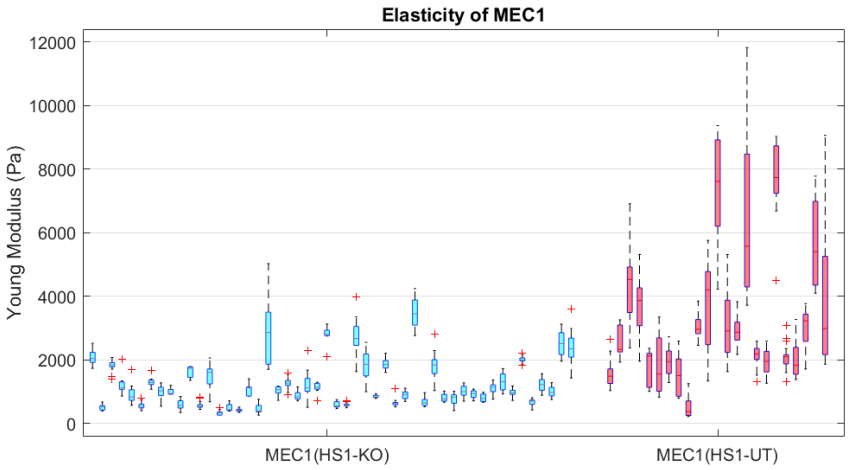

Figure B.7.: Young's modulus of elasticity calculated for $MEC1^{HS1KO}$ (blue) and $MEC1^{HS1UT}$ (pink) cells lines. Boxes represent the interquartile range; whiskers delimit the 25th and 75th percentile range, red crosses represents the outliers values.

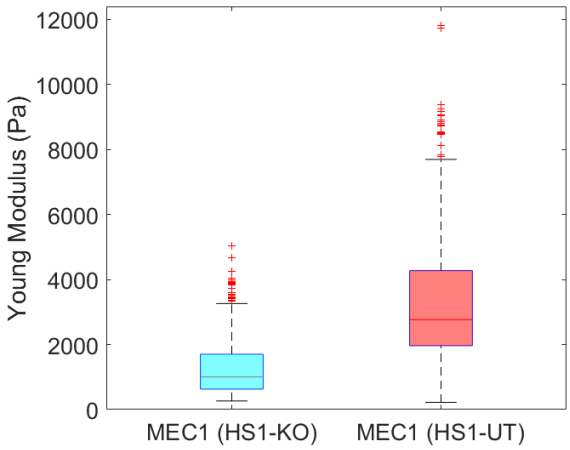

Figure B.8.: Merged Young's modulus of elasticity calculated for$MEC1^{HS1KO}$ (blue) and $MEC1^{HS1UT}$ (pink) cells lines. Boxes represent the interquartile range; whiskers delimit the 25th and 75th percentile range, red crosses represents the outliers values. Final Young Modulus values: $E_{MEC1}(HS1-KO) = 1.23 \pm 0.79$ kPa and $E_{MEC1}(HS1-UT) = 3.40 \pm 2.21$ kPa.

B.2.3 *Pancreatic ductal adenocarcinoma Cells*

Pancreatic Ductal Adenocarcinoma (PDAC) accounts for about the 90% of pancreatic cancers, characterized by a poor prognosis due to its particularly aggressive metastatic nature [289]. It has been previously reported that PDAC cells expresses high level of Gα16 protein, the α-subunit of heterotrimeric G16 protein [291]. In gastrointestinal and pancreatic tumors, the expression of this protein induces a neoplastic transformation of healthy cells.

This work had been focused on the study of the possible biological role of Gα16 in PT45 and PANC-1 human cell lines deriving from PDAC. Two different PT45 and PANC-1 cell line were generates, in which the expression of Gα16 was absent following gene silencing using the CRISPR/CAS9 technique (KO).

Previous studies showed that neoplastic cells might have a different membrane stiffness with respect to healthy cells [26, 27, 28, 290]. Thus, the contribution of Gα16 to the regulation of membrane stiffness was evaluated by and AFM-based single cell force spectroscopy.

The results showed that PANC-1 KO cells, which suffer from the absence of the Gα16 protein, were more rigid than their WT counterpart, indicating a possible involvement of Gα16 in the regulation of cellular stiffness (Fig.B.11). The same effect was not noticeable in PT45 cells (Fig.B.12).

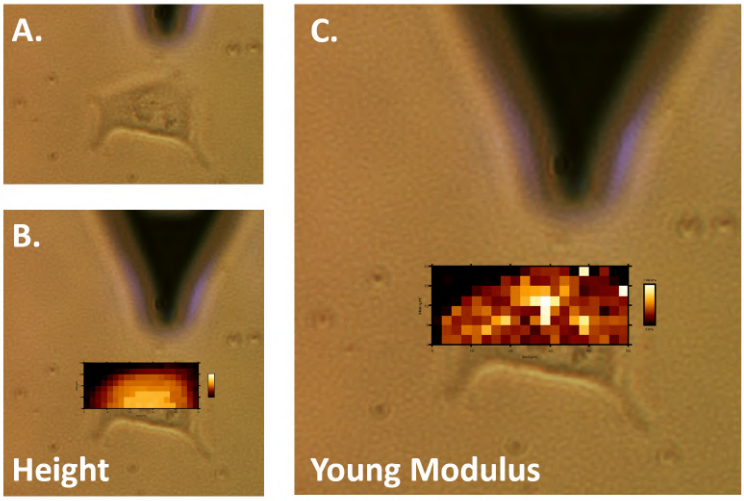

Figure B.9.: A typical experiment of PANC1-WT cell line elasticity. A) Optical image of a PANC1-WT cell in the field of view of AFM camera. B) AFM height image of the same cell (60 x 60 μm^2, 128 x 128 pixels, z-bar 14μm). C) Elasticity height image of the same cell (60 x 60 μm^2, 128 x 128 pixels, z-bar 8 kPa).

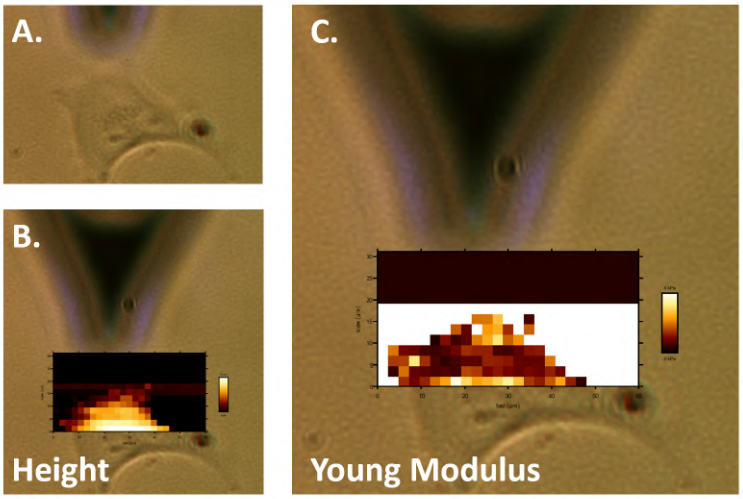

Figure B.10.: A typical experiment of PANC1-KO cell line elasticity. A) Optical image of a PANC1-KO cell in the field of view of AFM camera. B) AFM height image of the same cell (60 x 60 μm^2, 128 x 128 pixels, z-bar 14μm). C) Elasticity height image of the same cell (60 x 60 μm^2, 128 x 128 pixels, z-bar 8 kPa).

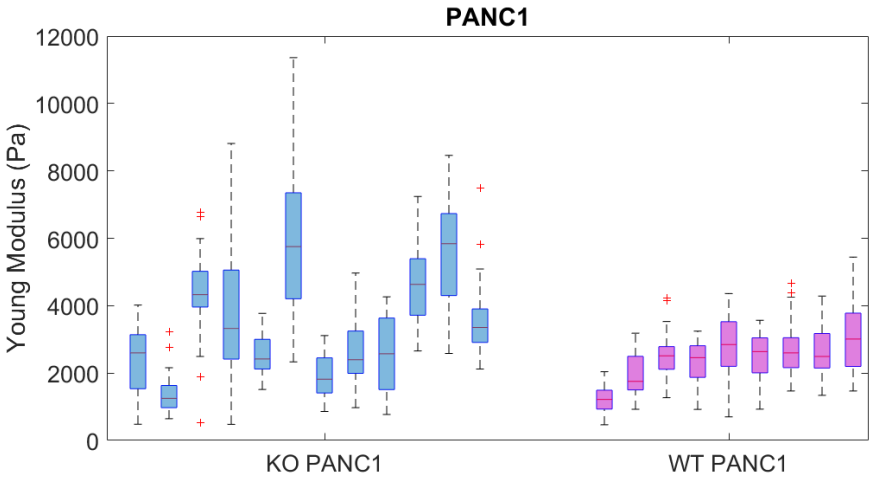

Figure B.11.: Young's modulus of elasticity calculated for PANC1-KO (blue) and PANC1-WT (red) cells lines. Boxes represent the interquartile range; whiskers delimit the 25th and 75th percentile range, red crosses represents the outliers values. Final Young Modulus values: $E(PANC1^{KO}) = 3.47 \pm 1.46$ kPa and $E(PANC1^{WT}) = 2.42 \pm 0.56$ kPa.

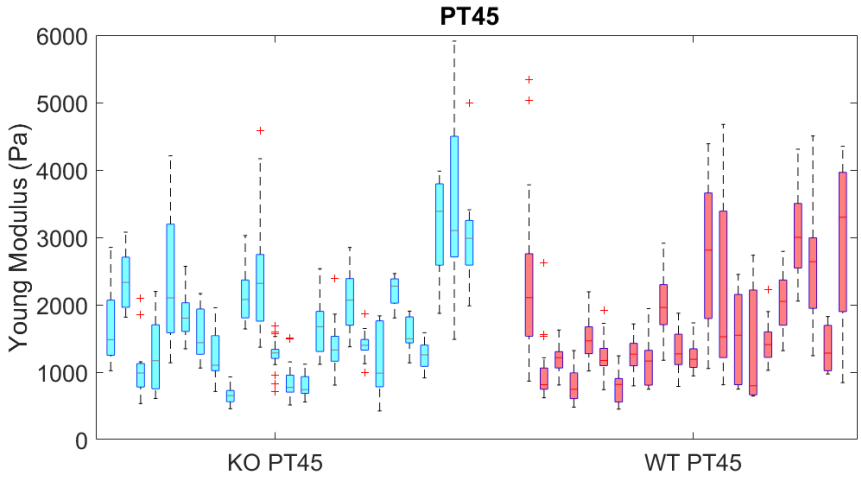

Figure B.12.: Young's modulus of elasticity calculated for PT45-KO (light blue) and PT45-WT (pink) cells lines. Boxes represent the interquartile range; whiskers delimit the 25th and 75th percentile range, red crosses represents the outliers values. Final Young Modulus values: $E(PT45^{KO}) = 1.76 \pm 0.75$ kPa and $E(PT45^{WT}) = 1.69 \pm 0.69$ kPa.

APPENDIX

C

α-SYN SUPPLEMENTARY

C.1 DNA PLASMID MAPS

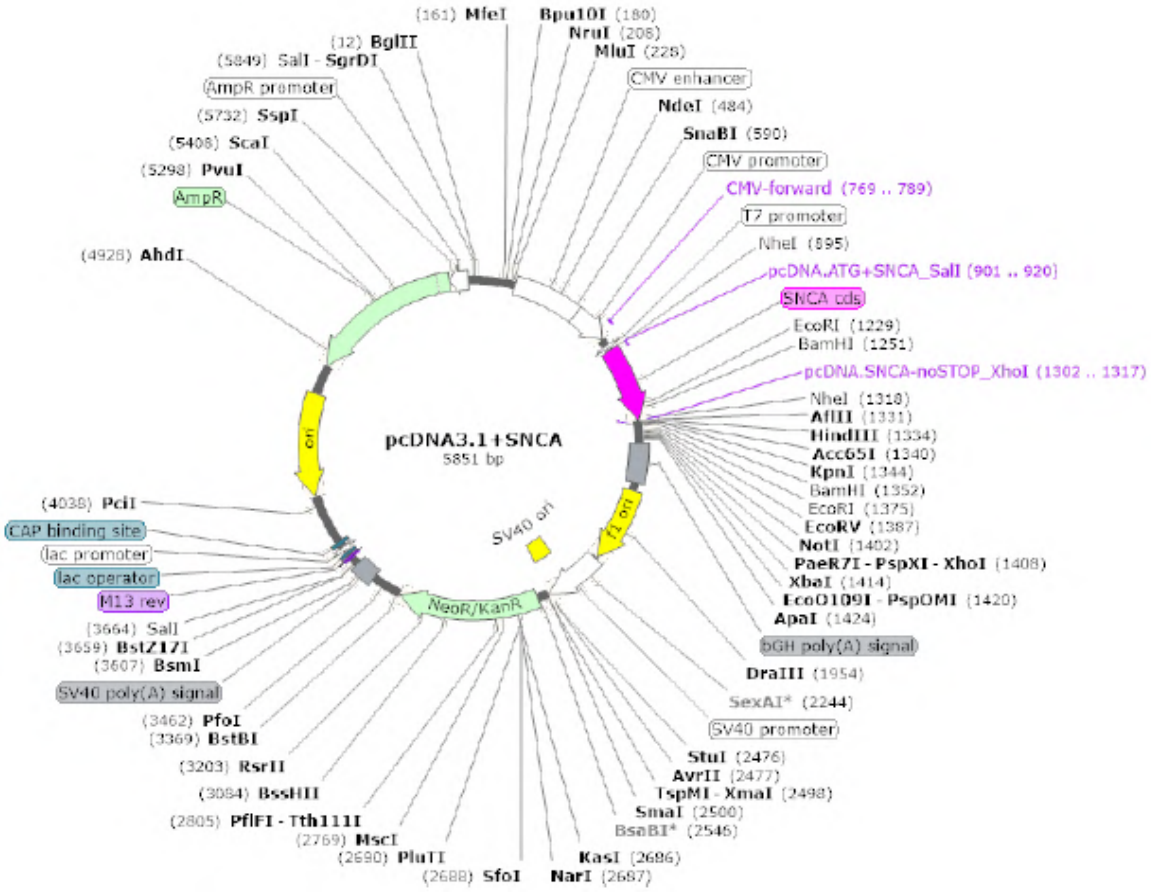

Figure C.1.: **Map of pcDNA 3.1 vector with the cDNA of α-syn WT**. The cDNA was cloned NheI oriented sense in the direction with respect to the T7 promoter. In violet the name and site of the primers for seqeuncing (CMV-forward) and cloning (pcDNA.ATG + SNCA-SalI, pcDNA.SNCA-noSTOP-XhoI) are indicated. In bold all the unique restriction sites are underlined.

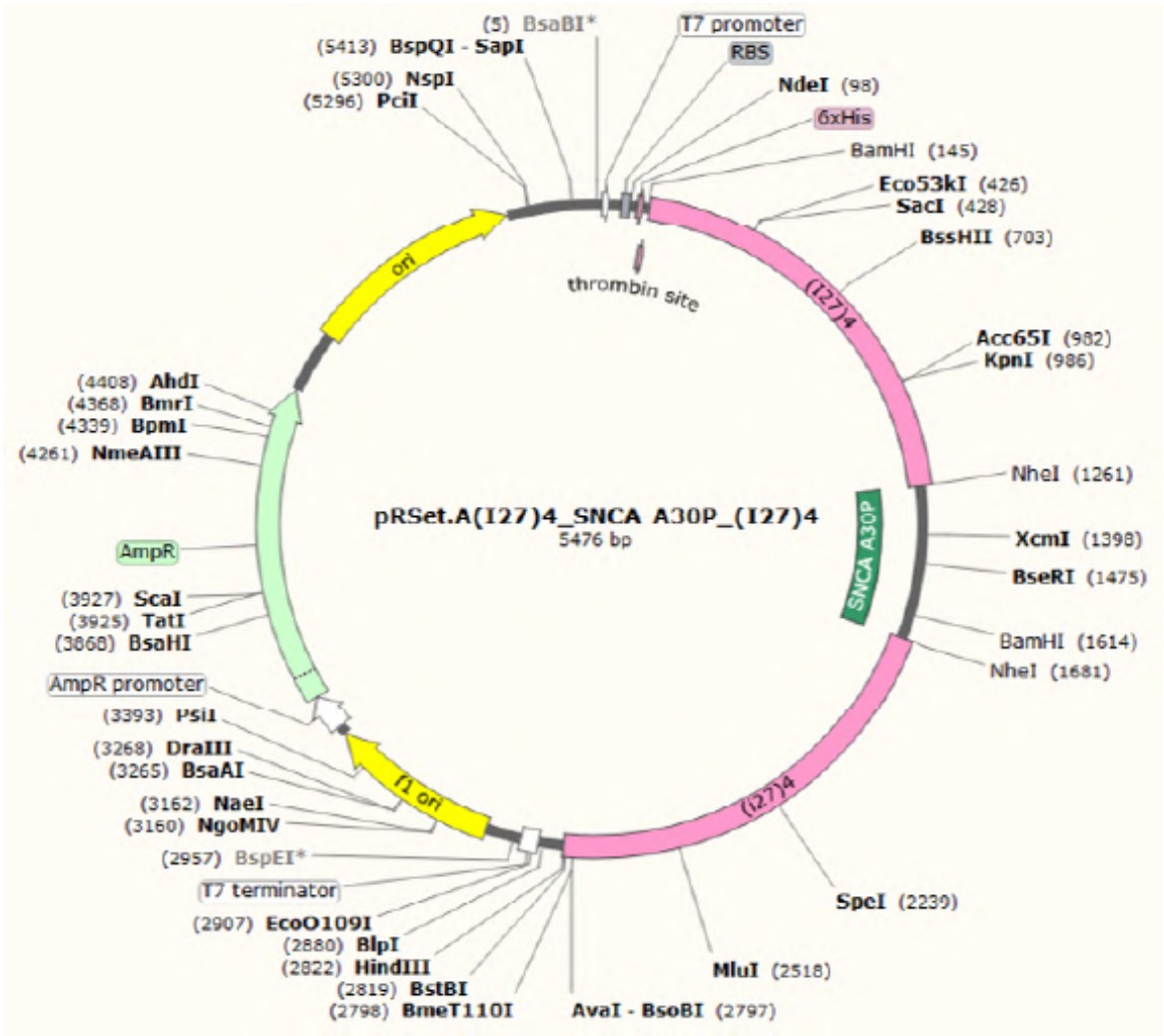

Figure C.2.: **Map of pRSetA $(I27)_4$-α-syn-$(I27)_4$ vector with SNCA fragment (example with A30P mutant)**. The T7 promoter, the SNCA A30P cDNA sequence (green) and the His-tag (6XHis), necessary for the protein purification, are shown in the picture. The NheI restriction site used for the mutant SNCA cDNA cloning is indicated upstream and downstream the sequence. In bold all the unique restriction sites are underlined.

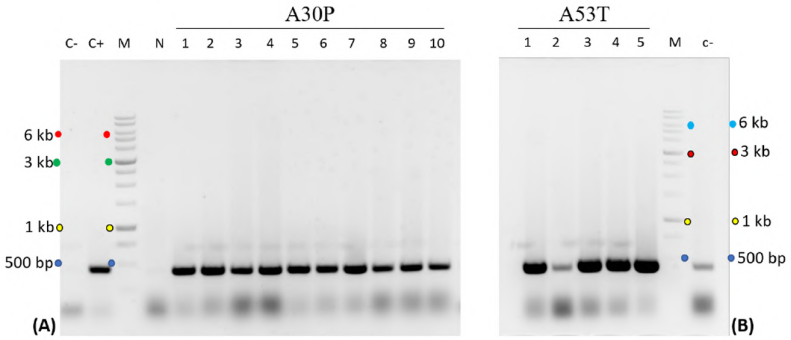

Figure C.3.: **Colony PCR for the screening of SNCA mutants in the pRSet.A**. 1% agarose gel, stained with ethidium bromide. M = 1kb DNA ladder (0.25 µg); the colour dots indicate the base pairs relative to every band. C- negative control (template: empty pRSet.A $(I27)_8$ vector), C+ positive control (template: pRSet.A $(I27)_4$- SNCA WT-$(I27)_4$), N negative control (template: a colony tranformed with the empty vector). A: 1-10 =A30P. B: 1-5 = A53T. For each sample 25 µL of PCR mixture were loaded.

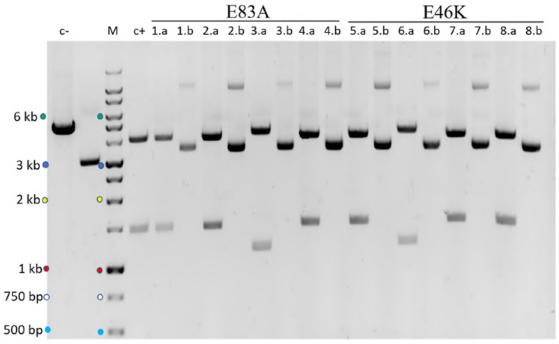

Figure C.4.: **Test digestion for the screening of the α-syn mutant direction of insertion in pRSet.A**. For each sample 200-250 ng of digested DNA were loaded on 1% agarose gel, stained with ethidium bromide. M = 1kb DNA ladder (0.25 µg). C- (template: a colony tranformed with the empty vector), C+ (template: pRSet.A $(I27)_4$-SNCA WT-$(I27)_4$). For all the clones also the non digested DNA was loaded on the gel: .a = non digested, .b = digested with BamHI. 1-4 = E83A, 5-8 = E46K.

D

IDEAL CHAINS MODELS FOR POLYMERS

Ideal chains are polymeric chains with no interaction (repulsive or attractive) among the n monomers by which they are composed and they follow a typical random walk. Real chains have not a perfect ideal behaviour, but there are, also in biology, polymeric systems which are well described by ideal chains-based models.

To describe the behaviour of a chain, we can consider the end to end distance $\langle R \rangle$. If the displacement of monomers has random nature, such as there is nor attraction or repulsion between them, the mean end to end distance is null $\langle R \rangle = 0$. The first non-trivial moment of the distribution of end-end distance is the second moment, so it is necessary to consider the mean squared end-end distance, $\langle R^2 \rangle$ defined as

$$\begin{aligned} \langle R^2 \rangle =& \langle \bar{R_n} \cdot \bar{R_n} \rangle \\ &= \left\langle (\sum_{i=1}^{n} \bar{r_i}) \cdot (\sum_{j=1}^{n} \bar{r_j}) \right\rangle \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \langle \bar{r_i} \cdot \bar{r_j} \rangle \\ &= l^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \langle \cos\vartheta_{ij} \rangle \end{aligned} \tag{D.1}$$

where ϑ is the bond angle between two consecutive monomers, l is the length of each monomer, $\bar{r_i}$ is the vector that links two consecutive monomers and $\bar{R_n}$ is the sum of all $\bar{r_i}$ of the chain.

This parameter has a different analytic expression, depending of the assumptions that are made on torsion angles permitted for each model.

D.1 FREELY JOINTED CHAIN MODEL

One of the simplest model for an ideal polymer is the Freely Jointed Chain (FJC) model, which has a constant bond length $l = |\mathbf{r}_i|$ and no correlation is considered between the direction of bond angles and no assumptions are made for torsional angles. In this case $\langle \cos\vartheta_{ij} \rangle = 0$ for $i \neq j$.

Then, the second moment of R is:

$$\langle R^2 \rangle = n_k l_k^2 \tag{D.2}$$

where n_k is the number of Kuhn segments and the l_k is the Kuhn length. The analytic expression which describe the end-to-end length x when an external force F is applied to the polymer is:

$$x(F) = L_c \left[\coth\left(\frac{F l_k}{k_B T}\right) - \frac{k_B T}{F l_k} \right] \tag{D.3}$$

The chain link size (Kuhn length) is l_k and the contour length $L_c = n_k \cdot l_k = n \cdot l$.

D.2 FREELY ROTATING CHAIN MODEL

In this model, the differences between the probabilities of different torsion angles is neglected and it is assumed all torsion angles to be equally probable. Thus, the Freely Rotating Chain (FRC) model ignores the variation of the potential due to the torsion angle and assumes all bond lengths and bond angles are constant.

The mean square end to end distance of a FRC can be written in terms of cosines:

$$\begin{aligned}
\langle R^2 \rangle &= \sum_{i=1}^{n} \sum_{j=1}^{n} \langle \bar{r_i} \cdot \bar{r_j} \rangle = \sum_{i=1}^{n} \left(\sum_{j=1}^{i-1} \langle \bar{r_i} \cdot \bar{r_j} \rangle + \langle \bar{r_i^2} \rangle + \sum_{j=i+1}^{n} \langle \bar{r_i} \cdot \bar{r_j} \rangle \right) \\
&= \sum_{i=1}^{n} \langle \bar{r_i^2} \rangle + l^2 \sum_{i=1}^{n} \left(\sum_{j=1}^{i-1} (\cos\vartheta)^{i-j} + \sum_{j=i+1}^{n} (\cos\vartheta)^{j-i} \right) \\
&= n l^2 + l^2 \sum_{i=1}^{n} \left(\sum_{k=1}^{i-1} \cos^k\vartheta + \sum_{k=1}^{n-i} \cos^k\vartheta \right)
\end{aligned} \tag{D.4}$$

Considering $(\cos\vartheta)^{|j-i|} = exp[-\frac{|j-i|}{s_p}]$, the segment of persistence length of the chain s_p[6] is

$$s_p = -\frac{1}{\ln(\cos\vartheta)} \tag{D.5}$$

The mean square end to end distance is then:

$$\langle R^2 \rangle = n l^2 + 2 n l^2 \frac{\cos\vartheta}{1 - \cos\vartheta} = n l^2 \frac{1 + \cos\vartheta}{1 - \cos\vartheta} \tag{D.6}$$

and the the Flory's characteristic rapport[7] for FRC (C_∞) is then

$$C_\infty = \frac{1 + \cos\vartheta}{1 - \cos\vartheta} \tag{D.7}$$

6 s_p represents the number of bonds which belong to this segment as long as the persistence length

7 considering $C'_i = \sum_{j=1}^{n} \langle \cos\vartheta_{ij} \rangle$ the value $C'_i \equiv C_\infty$ is the theoretical value for an infinite chain and depends on the stiffness of the polymer

The Worm Like Chain (WLC) model, somewhere also known as Kratky-Porod model, is a special case of the FRC model for small values of the bond angle and it is particularly suitable for describing stiff polymers, such as double strand DNA [292].

The bond angle ϑ between two consecutive monomers is very small ($\vartheta \ll 1$) and it is possible to make the following approximations for $\cos\vartheta$ and $\ln(\cos\vartheta)$:

$$\begin{aligned} \cos\vartheta &\simeq 1 - \frac{\vartheta^2}{2} \\ \ln(\cos\vartheta) &\simeq \frac{\vartheta^2}{2} \end{aligned} \tag{D.8}$$

The segment of persistence length of the chain (s_p), the Flory's characteristic rapport for WLC (C_∞) and the Kuhn length[8] (l_k) are:

$$\begin{aligned} s_p &\simeq \frac{2}{\vartheta^2} \\ C_\infty &\simeq \frac{4}{\vartheta^2} \\ l_k &\simeq l\frac{4}{\vartheta^2} = 2L_p \end{aligned} \tag{D.9}$$

The $\langle R^2 \rangle$ for 3-dimensional chains is:

$$\langle R^2 \rangle = 2L_p R_{max}\left(1 - \frac{L_p}{R_{max}}(1 - e^{-\frac{L_p}{R_{max}}})\right) \tag{D.10}$$

The main assumptions and predictions of ideal chains models (FJC, FRC and WLC) are reported in Tab.D.1.

Models	**FJC**	**FRC**	**WLC**
Bond length (l)	Fixed	Fixed	Fixed
Bond angle (ϑ)	Free	Fixed	Fixed ($\vartheta \to 0$)
Torsion angle (ϕ)	Free	Free	Free
Next ϕ independent?	Yes	Yes	Yes
C_∞	1	$\frac{1+\cos\vartheta}{1-\cos\vartheta}$	$\simeq \frac{4}{\vartheta^2}$

Table D.1.: Comparison of some ideal chain models. Adapted from [292]

8 l_k is the length of each segment on which the chain is divided. At the effective l_k the joints become free

D.3.1 *Stretching a Worm Like Chain polymer*

At finite temperatures, the distance between the two ends of the polymer will be significantly shorter than the contour length L_c[9]. This is caused by thermal fluctuations, which result in a coiled, random configuration of the polymer. Upon stretching the polymer, the accessible spectrum of fluctuations reduces, which causes an entropic force against the external elongation. This entropic force can be estimated by considering the entropic Hamiltonian:

$$H = H_{entropic} + H_{external} = \frac{1}{2} k_B T \int_0^{L_C} L_p \cdot \left(\frac{\partial^2 \bar{r}(s)}{\partial s^2} \right) ds - xF \quad \text{(D.11)}$$

where F is the external force that is applied to the polymer.

The outcome of there calculations can be summarized in an interpolation formula [293] of force (F) versus position (x):

$$F(x) = \frac{k_B T}{L_p} \left[\frac{1}{4} \left(1 - \frac{x}{L_c} \right)^{-2} + \frac{x}{L_c} - \frac{1}{4} \right] \quad \text{(D.12)}$$

9 The contour length is indicated as L_C in Part ii and L_0 in Part 11.1

E

STEP FINDING ALGORITHM: TECHNICAL CALIBRATION

E.1 STEP-FINDING-ALGORITHM VALIDATION

E.1.1 *Well defined steps (> 100nm) finding - 1 cycle iteration*

In order to confirm that the software implemented in MATLAB based on [252], an artificial step-like signal is acquired at the same force used for the condensin mediate DNA collapsing experiments (F = 0.5pN) for a single Nicked DNA of ~ 6.2kpb.

To obtain a step-like function, the objective-linked piezo was manually displaced vertically in both the directions (up and down) by steps of 100, 200 and 1000 nm of variable time duration, resulting in a well known step-like signal (Fig. E.1) with a true thermal noise (of about $\sigma \sim 55nm$). In this preliminary code the *counter* fit was rigidly shifted by 250-500 points.

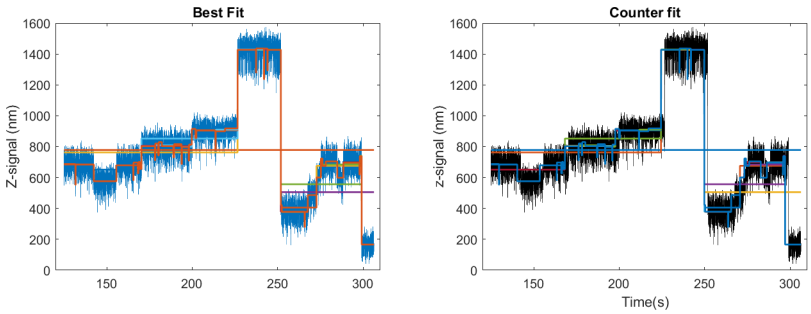

Figure E.1.: Data trace with best fit (for N=1:50) and the counter fit shift (shift=250 points)

For the trace shown here ($N_{bestfit} = 12$), we obtain a $\chi^2 = 6.37 \cdot 10^7$ which is completely in agrement with the expected value $\chi^2_{GoodFit} \sim N_{sample} \cdot \sigma^2 = 6.50 \cdot 10^7$.

E.1.2 *Well defined steps (> 100nm) finding - 2 cycle iterations*

For particularly complicated traces, like very long traces and with a noise comparable to the height of the steps, it has been demonstrated by Eeftens and colleagues [254] that repeating the algorithm twice

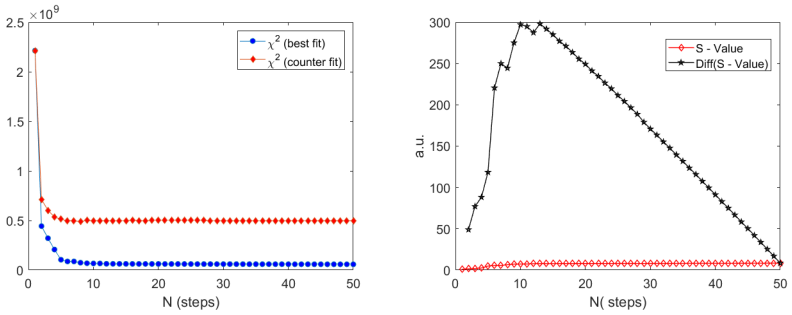

Figure E.2.: χ^2 (left) and S-value and 1^{st} derivative of S-value (right)

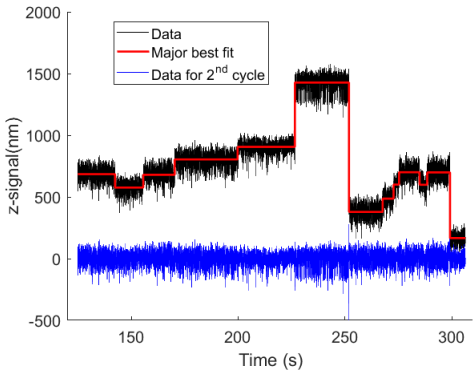

Figure E.3.: Final N(step)=12 fit (red line) and the difference between signal and step-fit, highlighting that there is no missed step

could increase the goodness of the fit. This process was easily implemented into the code, by a adding a second cycle considering as data the difference between the raw data and the 1^{st} (or Major) fit.

To assess if the fit should converge also in the presence of a 2-step iteration, we use the same routine as the single cycle algorithm, by artificially lowing the $N_{bestfit}$ from 12 to 8 (Fig.E.5). In the real condensin:DNA data this second-refining routine was particularly important for a S-value signal with multiple peaks.

Step-finding-Algorithm validation: 10 traces

A total of n=10 DNA mixed (TC, Nicked and Torsionally constrained) traces (simultaneously acquired) are analysed by means of the 1-cycle and 2-cycled algorithm to validate the code. For the 1-cycle Step-finding-Algorithm the right number of steps is found 7/10 times, and the found steps hight are in well agreement with the expected ones ($\Delta Y = \pm 100, 500$ or $1000nm$, Tab.E.1).

For the two iteration cycle FSA, all the traces found at least 12 steps (the correct N) but some of them (4/10) find additional smaller steps ($\Delta Y \sim 20 : 50nm$). The steps follows a random distribution and are correlated with a more noisy experiment (Fig. E.6).

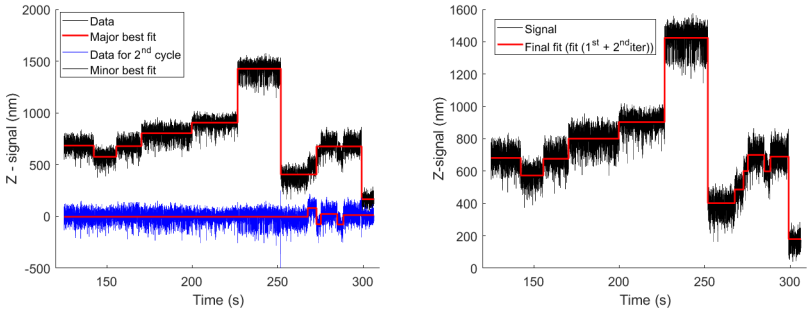

Figure E.4.: Two iteration fitting routine. Major best fit (N=8), Minor best fit (N=5). The sum has N=12 and it is not distinguishable from the one-iteration procedure

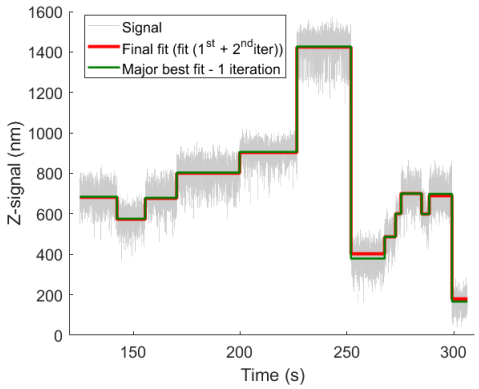

Figure E.5.: Comparison between 1 (green line) and 2 iteration (red line). The fitted N is exactly the same, although there is a slight difference in the y position of some of the founded steps

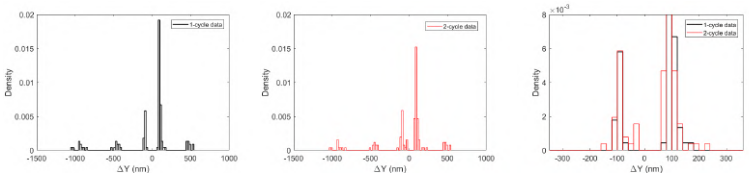

Figure E.6.: Resulting steps size for 10 traces by using a single iteration (left, black) or a double iteration process (center, red). The smaller steps (namely, between -100 and +100 nm) are more frequent for the 2-steps cycle algorithm.

E.1.3 *Implementation on FSA*

A new implementation of the Finding-Steps-Algorithm (FSA) has been done: basically,the *counter fit* is built up as a step-like function with the steps placed exactly in the middle of the steps find with the *best fit*.

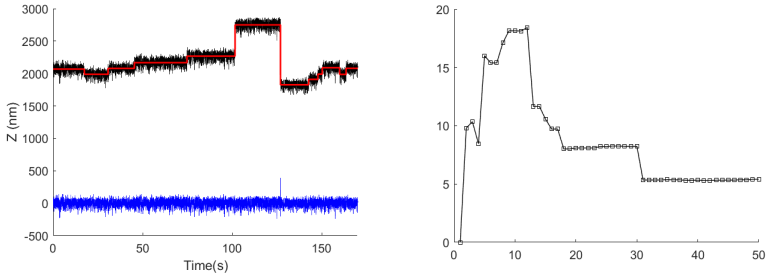

Figure E.7.: Left: Raw data (black) and the 1^{st} iteration cycle N_{1st}(step)=5 fit (red line) and the difference between signal and step-fit (blue), highlighting that there are no steps missing. Right: The (S) Step-function for the 1st cycle of iteration.

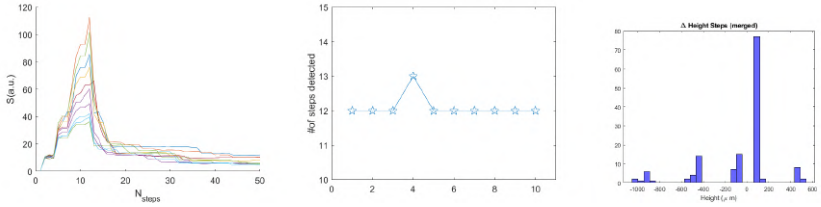

Figure E.8.: Left: number of N_{steps} find. Middle: The Step-function for the 1st cycle of iteration for all the 10 beads Left: The step-heigth values found for the 10 traces.

E.2 STEP-FINDING-ALGORITHM-OPTIMIZED

Well defined steps ($>$ 100nm) finding

In order to confirm that the software gifted by Salapaka et al based on [253], I analysed the same artificial step-like signal as used in the previous sections.

With respect to the previous algorithm, two parameters have been fully analysed: the Likelihood-Ratio (LLR) parameter and the *measnoise* parameter.

While the LLR values are a result of the fitting procedure, the *measnoise* parameter can be initialised and is, basically, an estimation of the thermal noise of a trace (here, it is always intended in nm).

In this analysis the found LLRs values for all of them are clearly above the threshold (LLR > 1.5) and no *measnoise* was initialised.

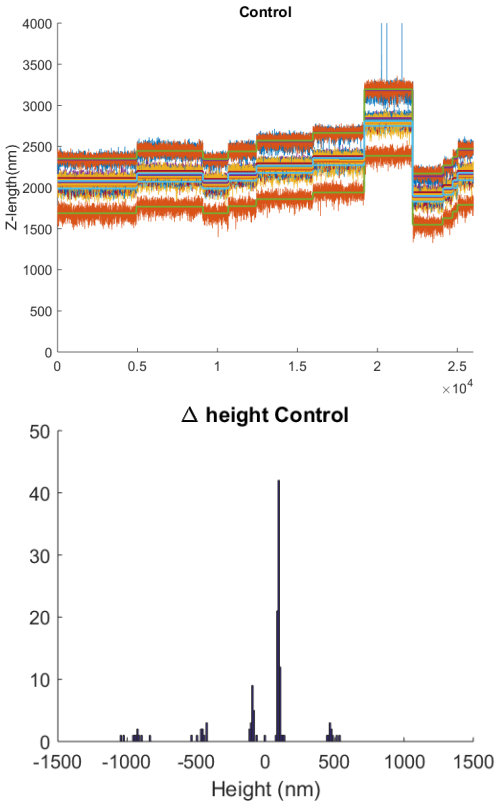

Figure E.9.: Top: The 10 DNA traces fitted by Step-Finding-Algorithm-optimized (SFAO)
Bottom: The ΔHeight e of the steps.

Steps found by using the Step-Finding-Algorithm

Real Steps (nm)	-1000	-500	-100	100	500
SFA* (1-iter)	-942 ± 60	-457 ± 35	-83 ± 27	98 ± 26	478 ± 28
SFA* (2-iter)	-922 ± 58	-446 ± 31	-92 ± 10	99 ± 15	479 ± 26
SFA° (2-iter)	-941 ± 60	-459 ± 31	-89 ± 11	94 ± 9	479 ± 26
SFAO	-941 ± 60	-458 ± 35	-94 ± 12	95 ± 10	478 ± 27

Table E.1.: (*)The height of the steps found by Step-Finding-Algorithm for 1(1^{st} line) or 2(2^{nd} line) cycle iterations wih a counter fit shifted by 250 points. (°)In the third line a 2-iter algorithm with the counter fit step-like function had the steps placed exactly in the middle of the steps found in the first iteration. In the last line are shown the steps found with the Step-Finding-Algorithm-optimized (SFAO).

The intrinsically thermal noise of an experiment is strictly linked to non-controllable factors (as the temperature or the length of DNA) or to instrumental-based errors (rigid drifting during time or precision of the bead-tracking algorithm, as well as the reproducibility of the data). All of these factors influences the reliability of the steps found by both SFA and SFAO.

To characterized the intrinsically stability of the instrumental setup, several trials has been performed in the same condition of the experiments:

- Total DNA tracked: 6.
- Standard Flow Cells, 1 μm beads, DNA G12E, 1mm gap magnets.
- Forces considered: 1pN, 0.5 pN, 0.3 pN.
- Stepsize considered: 100, 50, 30 and 20 nm.
- Algorithm employed: SFAO.
- *measnoise* considered: free, 30, 50, 100, 200.

These experiments have been performed by me and Dr. Julene Madriarga-Marcos, based on an implementation of the acquiring software done by Maite Arranz.

Force = 1 pN

Table E.2.: **StepSize 100 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-88.7 ± 23.0	75.4 ± 30.7	6/6
noise 30	-85.2 ± 26.5	67.7 ± 20.3	6/6
noise 50	-84.2 ± 23.8	70.2 ± 21.5	6/6
noise 100	-93.5 ± 15.3	73.45 ± 15.0	6/6
noise 200	-144.9 ± 12.5	132.7 ± 33.0	0/6

Table E.3.: **StepSize 50 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-49.2 ± 14.0	36.0 ± 12.1	4/4
noise 30	-47.45 ± 17.1	36.8 ± 10.4	4/4
noise 50	-58.8 ± 11.8	41.9 ± 12.7	4/4
noise 100	-77.57 ± 10.8	62.4 ± 13.7	3/4
noise 200	-115.5 ± 27.0	101.3 ± 33.5	4/4

Table E.4.: **StepSize 30 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-36.6 ± 9.9	20.8 ± 9.8	0/6
noise 30	-37.5 ± 8.8	21.8 ± 14.5	1/6
noise 50	-40.0 ± 6.1	28.2 ± 8.8	0/6
noise 100	-56.9 ± 11.6	44.1 ± 9.7	0/6
noise 200	-89.9 ± 14.2	75.7 ± 13.9	4/6

Table E.5.: **StepSize 20 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-28.48 ± 9.6	16 ± 5.4	1/6
noise 30	-29.27 ± 9.8	14.5 ± 7.8	1/6
noise 50	-38.0 ± 8.6	23.4 ± 7.9	0/6
noise 100	-57.8 ± 7.0	47.8 ± 11.5	0/6
noise 200	-88.2 ± 12.7	81.18 ± 14.9	2/6

Double Gaussian Fit. Steps found by SFAO, 10 iterations. Histograms [-200:200] nm, bin 15nm.

Force = 0.5 pN

Table E.6.: **StepSize 100 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-86.2 ± 24.2	65.7 ± 8.2	3/3
noise 30	-75.6 ± 17.5	62.4 ± 10.4	3/3
noise 50	-86.7 ± 20.0	73.2 ± 21.1	3/3
noise 100	-91.6 ± 9.5	97.5 ± 12.8	3/3
noise 200	-160.6 ± 29.7	151 ± 54.9	0/3

Table E.7.: **StepSize 50 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-62.5 ± 26.3	45.8 ± 15.7	4/6
noise 30	-78.2 ± 16.9	68.9 ± 28.1	6/6
noise 50	-57.6 ± 19.7	45.2 ± 15.3	4/6
noise 100	-77.38 ± 26.9	60.6 ± 14.4	0/6
noise 200	-126.8 ± 11.5	108 ± 13.7	0/6

Table E.8.: **StepSize 30 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-68.3 ± 34.3	52.6 ± 36.7	4/6
noise 30	-105.6 ± 7.1	70.7 ± 12.2	6/6
noise 50	-73.6 ± 38.0	65.6 ± 41.9	4/6
noise 100	-68.4 ± 8.9	53.9 ± 12.6	0/6
noise 200	-118.7 ± 40.7	97.4 ± 25.1	0/6

Table E.9.: **StepSize 20 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-68.9 ± 54.5	–	4/6
noise 30	-86.1 ± 128.1	73.4 ± 23.2	6/6
noise 50	-74.7 ± 49	66.3 ± 44.7	4/6
noise 100	-60.23 ± 11.7	48.07 ± 14.4	0/6
noise 200	-97.7 ± 13.1	81.5 ± 13.5	0/6

Double Gaussian Fit. Steps found by SFAO, 10 iterations. Histograms [-200:200] nm, bin 15 nm.

Force = 0.3 pN

Table E.10.: **StepSize 100 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-118.9 ± 51.6	97.6 ± 25.7	3/3
noise 30	-114.3 ± 15.8	95.9 ± 17.3	3/3
noise 50	-100.3 ± 9.7	104.2 ± 14.3	3/3
noise 100	-118.0 ± 27.0	88.16 ± 10.0	3/3
noise 200	-172.8 ± 36.4	133 ± 21.2	2/3

Table E.11.: **StepSize 50 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-104 ± 29.0	100.6 ± 50.4	3/3
noise 30	-112.4 ± 27.8	97.8 ± 15.8	3/3
noise 50	-108.9 ± 10.8	79.8 ± 8.9	3/3
noise 100	-89.9 ± 13.9	82.8 ± 10.5	2/3
noise 200	-134 ± 7.0	107.8 ± 12.3	0/3

Table E.12.: **StepSize 30 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-100.4 ± 33.4	109.3 ± 7.3	3/3
noise 30	-107.9 ± 25.23	95.1 ± 37.0	3/3
noise 50	-114.2 ± 25.2	95.4 ± 12.7	3/3
noise 100	-79.6 ± 27.1	66.9 ± 20.4	1/3
noise 200	-121 ± 7.0	91.6 ± 15.5	0/3

Table E.13.: **StepSize 20 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-109.5 ± 51.3	94.3 ± 41.6	3/3
noise 30	-98.3 ± 26.9	101.8 ± 35.0	3/3
noise 50	-117.6 ± 42.8	110.2 ± 6.8	3/3
noise 100	-97.85 ± 32.8	50.28 ± 10.97	1/3
noise 200	-101.8 ± 14.6	94.0 ± 45.34	0/3

Double Gaussian Fit. Steps found by SFAO, 10 iterations. Histograms [-200:200] nm, bin 15 nm.

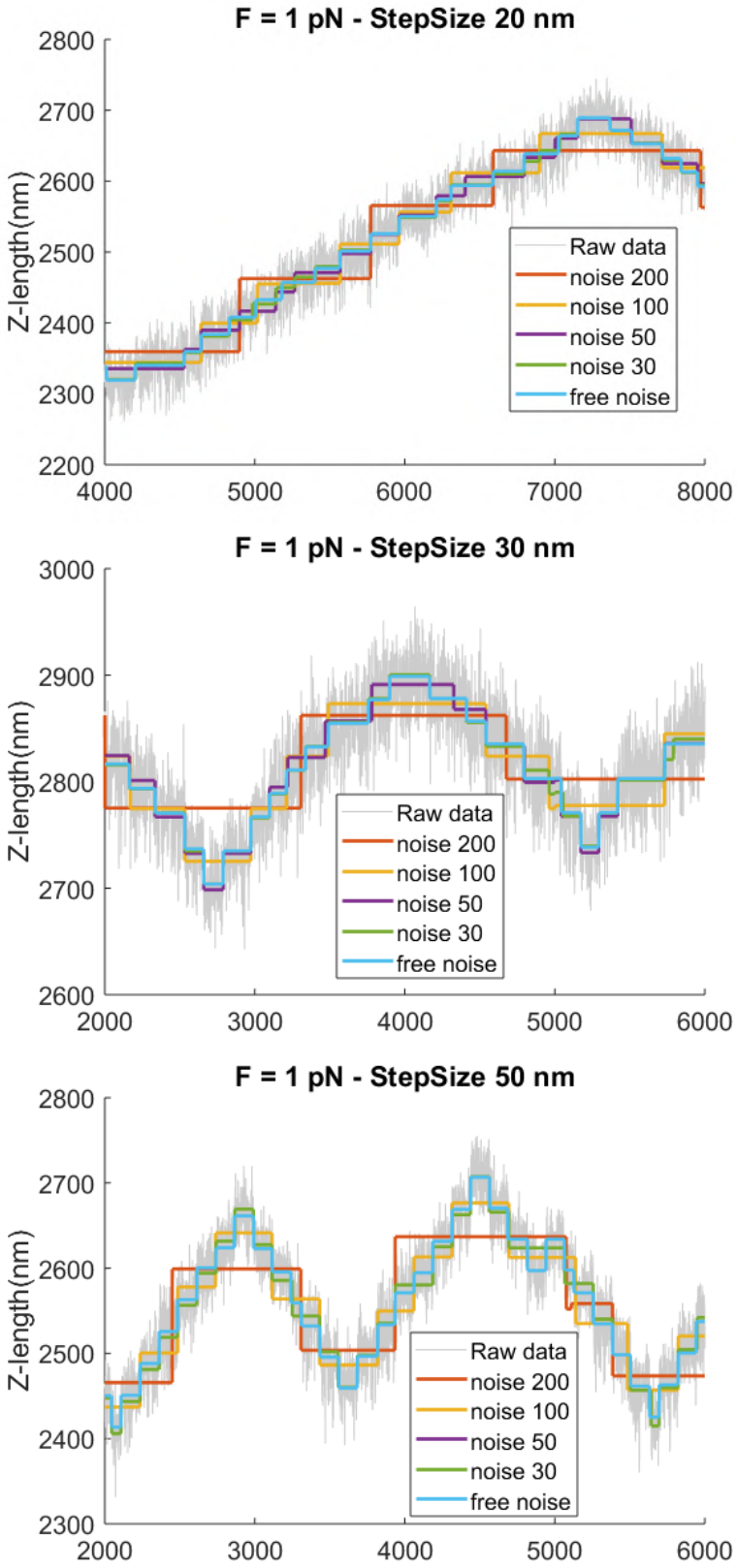

Figure E.10.: Representative trace of a DNA at different StepSize (Force 1pN) and their step trace calculated by means of SFAO (measnoise 200, 100, 50, 30, free). Length in x 4000 points/33 s.

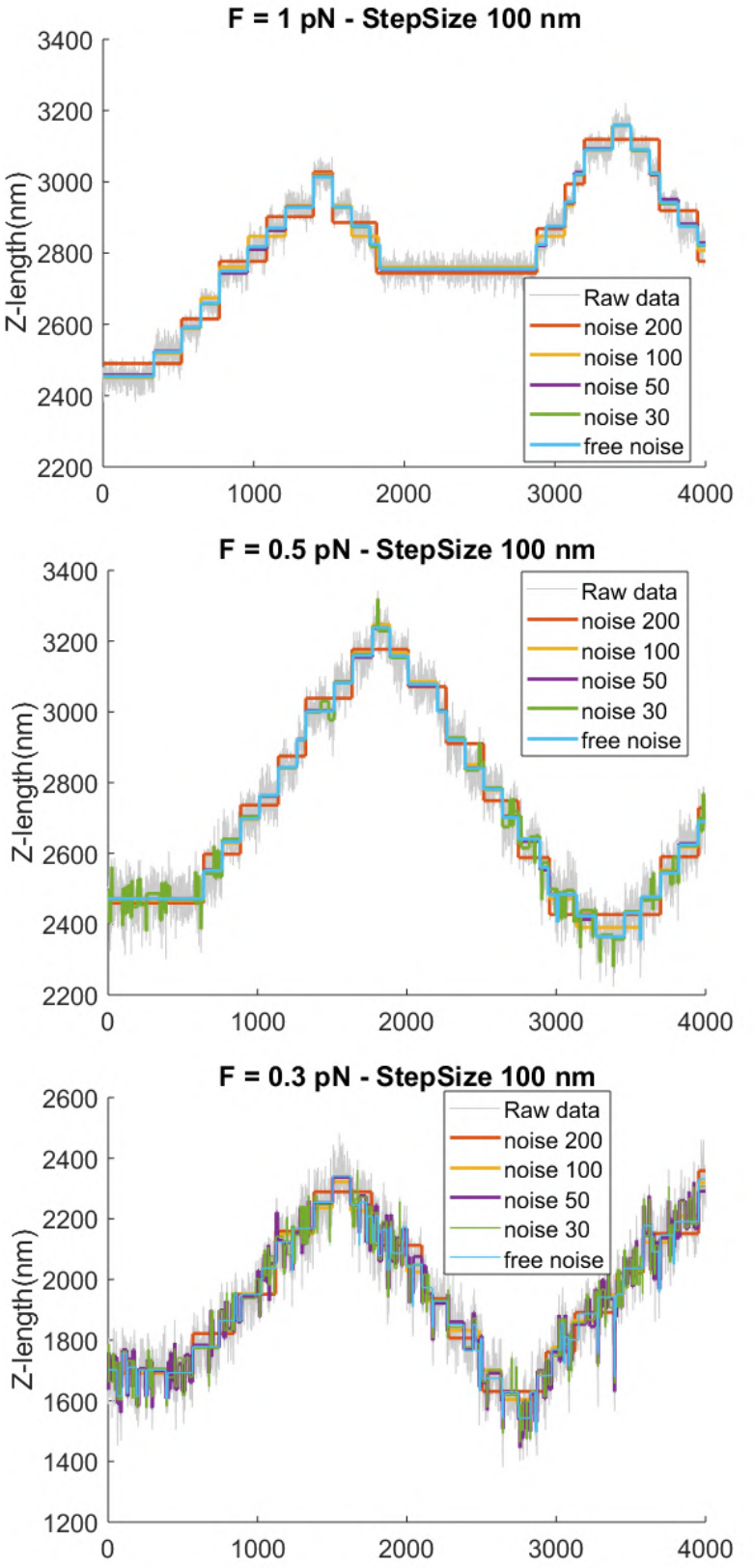

Figure E.11.: Representative trace of a DNA at Different Forces (StepSize = 100nm) and their step trace calculated by means of SFAO (measnoise 200, 100, 50, 30, free). Length in x 4000 points/33 s.

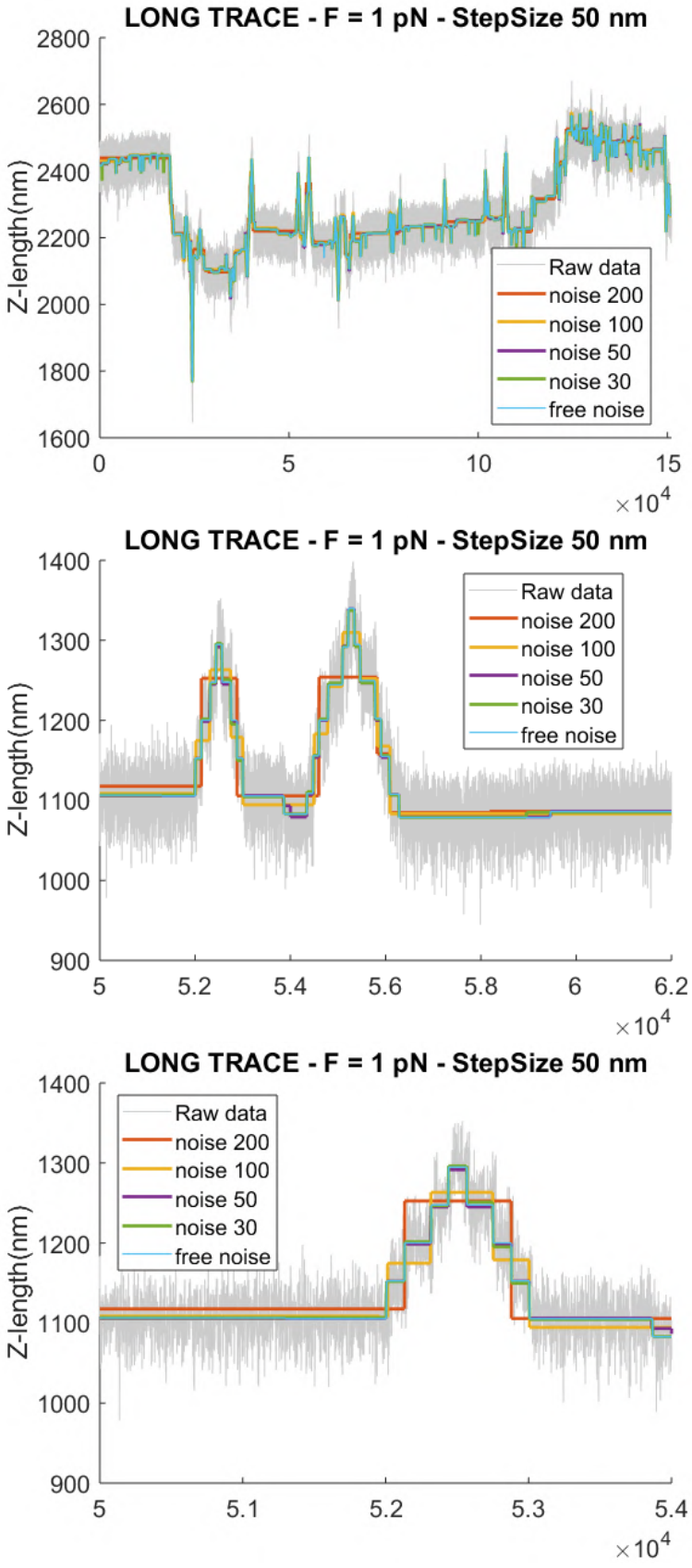

Figure E.12.: Representative trace of a DNA at F=1pN(StepSize = 50nm) and their step trace calculated by means of SFAO (measnoise 200, 100, 50, 30, free). Length in x ~ 21 min (Top), ~ 90 s (Middle) and 33 s (Bottom).

Force = 1 pN - Short trace vs Long Trace

Table E.14.: **SHORT TRACE StepSize 50 nm - Total Time ~ 3 min**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-47.9 ± 14.3	38.4 ± 14.3	4/4
noise 30	-48.6 ± 14.8	38.9 ± 7.3	4/4
noise 50	-55.7 ± 6.5	44.6 ± 10.3	4/4
noise 100	-78.1 ± 6.1	64.9 ± 7.1	3/4
noise 200	-114.9 ± 34.6	104.4 ± 32.7	4/4

Table E.15.: **LONG TRACE: StepSize 50 nm - Total Time ~ 22 min**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-48.9 ± 7.2	39.8 ± 7.2	3/3
noise 30	-48.7 ± 8.3	38.4 ± 10.8	3/3
noise 50	-49.7 ± 7.6	42.6 ± 8.8	3/3
noise 100	-70.8 ± 39.4	56.8 ± 7.6	3/3
noise 200	-89.9 ± 46.5	95.0 ± 56.1	3/3

Double Gaussian Fit. Steps found by SFAO, 10 iterations. Histograms [-200:200] nm, bin 10 nm.

SMOOTH 10 Force = 1 pN

Table E.16.: **StepSize 100 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-62.4 ± 35.5	48.9 ± 40.2	6/6
noise 30	-82.10 ± 25.2	69.1 ± 22.6	6/6
noise 50	-85.2 ± 23.5	70.2 ± 13.0	6/6
noise 100	-85.9 ± 25.0	68.6 ± 9.2	6/6
noise 200	-78.6 ± 10.1	65.5 ± 6.2	0/6

Table E.17.: **StepSize 50 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-37.3 ± 17.0	$19.4.0 \pm 19.9$	4/4
noise 30	-49.7 ± 14.9	35.8 ± 11.4	4/4
noise 50	-53.0 ± 12.1	37.5 ± 11.2	4/4
noise 100	-56.2 ± 8.2	39.0 ± 10.3	4/4
noise 200	-56.2 ± 6.7	39.2 ± 9.9	4/4

Table E.18.: **StepSize 30 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-33.9 ± 10.6	20.6 ± 12.4	6/6
noise 30	-35.1 ± 9.7	19.4 ± 10.2	6/6
noise 50	-41.4 ± 9.0	27.6 ± 8.4	6/6
noise 100	-48.6 ± 12.9	30.3 ± 8.8	3/6
noise 200	-45.9 ± 20.0	32.3 ± 10.7	4/6

Table E.19.: **StepSize 20 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-28.26 ± 9.7	15.3 ± 5.2	6/6
noise 30	-28.6 ± 6.2	14.0 ± 8.7	6/6
noise 50	-39.8 ± 6.0	27.3 ± 8.3	4/6
noise 100	-44.1 ± 10.2	27.4 ± 7.8	3/6
noise 200	-46.1 ± 12.6	30.3 ± 6.4	1/6

Double Gaussian Fit. Steps found by SFAO, 10 iterations. Histograms [-200:200] nm, bin 15 nm.

SMOOTH 10 Force = 0.5 pN

Table E.20.: **StepSize 100 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-53.6 ± 8.4	40.9 ± 11.7	3/3
noise 30	-65.8 ± 25.5	56.1 ± 18.2	3/3
noise 50	-83.5 ± 21.1	72.3 ± 16.2	3/3
noise 100	-84.5 ± 9.4	71.8 ± 14.6	3/3
noise 200	-89.4 ± 22.6	80.6 ± 21.5	3/3

Table E.21.: **StepSize 50 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-55.7 ± 16.0	38.1 ± 23.4	6/6
noise 30	-57.0 ± 15.1	41.8 ± 13.1	6/6
noise 50	-55.8 ± 14.6	42.1 ± 14.1	6/6
noise 100	-76.4 ± 16.9	61.6 ± 13.3	3/6
noise 200	-81.6 ± 24.3	73.1 ± 21.4	1/6

Table E.22.: **StepSize 30 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-48.9 ± 10.0	39.7 ± 15.2	6/6
noise 30	-54.7 ± 17.5	43.1 ± 24.1	6/6
noise 50	-53.5 ± 8.4	35.0 ± 14.5	6/6
noise 100	-74.2 ± 22.3	44.1 ± 9.9	2/6
noise 200	-85.2 ± 9.4	73.1 ± 11.2	0/6

Table E.23.: **StepSize 20 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-49.3 ± 13.8	28.9 ± 11.7	6/6
noise 30	-51.2 ± 16.7	44.4 ± 18.7	6/6
noise 50	-51.0 ± 12.3	37.3 ± 17.0	6/6
noise 100	-67.4 ± 10.4	43.0 ± 10.5	0/6
noise 200	-83.3 ± 11.8	45.6 ± 11.1	0/6

Double Gaussian Fit. Steps found by SFAO, 10 iterations. Histograms [-200:200] nm, bin 15 nm.

SMOOTH 10 Force = 0.3 pN

Table E.24.: **StepSize 100 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-64.9 ± 10.2	59.5 ± 13.6	3/3
noise 30	-73.9 ± 17.5	59.0 ± 19.1	3/3
noise 50	-95.2 ± 33.9	72.7 ± 26.5	3/3
noise 100	-110.5 ± 18.1	86.4 ± 13.6	3/3
noise 200	-160.4 ± 7.2	93.5 ± 14.2	3/3

Table E.25.: **StepSize 50 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-67.1 ± 9.9	57.7 ± 24.9	3/3
noise 30	-76.0 ± 10.4	61.8 ± 20.1	3/3
noise 50	-77.7 ± 11.1	72.7 ± 30.3	3/3
noise 100	-94.6 ± 23.8	79.8 ± 13.2	3/3
noise 200	-130.2 ± 6.9	94.4 ± 20.7	2/3

Table E.26.: **StepSize 30 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-69.1 ± 7.1	49.0 ± 6.8	3/3
noise 30	-63.1 ± 7.1	53.3 ± 11.7	3/3
noise 50	-95.3 ± 9.9	83.3 ± 5.6	3/3
noise 100	-76.2 ± 25.9	62.0 ± 7.2	3/3
noise 200	-109 ± 43.5	77.4 ± 7.2	1/3

Table E.27.: **StepSize 20 nm**

	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok
free noise	-63.1 ± 21.5	46.8 ± 13.6	3/3
noise 30	-74.7 ± 11.1	53.1 ± 11.3	3/3
noise 50	-117.6 ± 42.8	110.2 ± 6.8	3/3
noise 100	-85.9 ± 16.7	48.76 ± 10.6	1/3
noise 200	-90.3 ± 6.7	86.3 ± 27.8	1/3

Double Gaussian Fit. Steps found by SFAO, 10 iterations. Histograms [-200:200] nm, bin 15nm.

E.4 DIFFERENT FOCAL POSITIONS

In a typical MT experiment, the height of a DNA molecule is derived from the x-y position of its tethered bead during time. Prior to each experiment, a Look-Up-Table is acquired for each bead, spanning ~ 4μm for a typical 2μm-long DNA. To assess the contribution of the different focal positions of the beads, a experimental session similar to the one reported in the previous section has been done.

- Total DNA tracked: 4
- Standard Flow Cells, 1 μm beads, DNA G12E, 1mm gap magnets.
- 3 different focal positions: f1 (~ 5μm from fully focused bead),f2 (~ 6μm)and f3 (~ 7μm).
- Forces considered: 1pN, 0.5 pN, 0.3 pN.
- Stepsize considered: 100, 50 and 30 nm.
- Algorithm employed: SFAO.
- *measnoise* considered: free, 30, 50, 100.
- *measnoise* considered (for data smoothed by a 10-points-gaussian-filter): free, 30, 50, 100.

Force = 1 pN

Table E.28.: **StepSize 100 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	Nº steps
free	f1	-105.5 ± 13.1	89.8 ± 10.5	3/3	414
	f2	-87.9 ± 46.6	91.4 ± 40.0	4/4	1226
	f3	-99.3 ± 32.3	81.4 ± 15.9	4/4	1132
30	f1	-91.3 ± 39.6	88.8 ± 13.7	3/3	579
	f2	-85.4 ± 46.7	107.8 ± 4.2	4/4	1785
	f3	-85.1 ± 37.1	98.1 ± 4.0	4/4	1831
50	f1	-103.7 ± 7.2	93.4pm5.1	3/3	356
	f2	-102.1 ± 16.0	94.5 ± 32.8	4/4	1053
	f3	-100.4 ± 34.6	87.1 ± 5.4	4/4	1124
100	f1	-105.1 ± 4.4	90.0 ± 2.9	3/3	358
	f2	-97.7 ± 12.8	83.5 ± 4.4	4/4	541
	f3	-95.6 ± 4.8	80.0 ± 3.2	4/4	421

Table E.29.: **StepSize 50 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	Nº steps
noise	f1	-56.4 ± 5.2	47.5 ± 5.4	3/3	510
	f2	-57.5 ± 15.3	43.2 ± 8.7	4/4	1203
	f3	-51.6 ± 10.2	59.5 ± 45.1	4/4	1121
30	f1	-56.7 ± 6.9	47.5 ± 7.3	3/3	598
	f2	-55.7 ± 8.2	47.1 ± 6.6	4/4	1539
	f3	-71.9 ± 37.5	65.1 ± 34.2	4/4	1573
50	f1	-60.8 ± 5.8	49.3 ± 6.9	3/3	439
	f2	-53.2 ± 7.6	41.1 ± 10.1	4/4	700
	f3	-53.6 ± 7.5	41.6 ± 10.0	4/4	823
100	f1	-33.6 ± 6.4	26.7 ± 7.0	0/3	20
	f2	-30 ± 8.45	40.0 ± 3.3	0/4	17
	f3	-22.6 ± 5.5	64.1 ± 5.1	1/4	16

Table E.30.: **StepSize 30 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	Nº steps
free	f1	-36.4 ± 4.6	28.1 ± 6.5	1/3	514
	f2	-41.2 ± 12.8	31.4 ± 17.8	3/4	890
	f3	-36.8 ± 13.5	41.6 ± 35.1	2/4	887
30	f1	-36.8 ± 7.7	28.3 ± 8.7	1/3	528
	f2	-75.0 ± 53.5	50.2 ± 67.5	3/4	1281
	f3	-80 ± 4.2	33.1 ± 20.9	2/4	1133
50	f1	-30.4 ± 7.0	19.4 ± 16.3	1/3	55
	f2	-88.6 ± 6.5	37.4 ± 8.3	1/4	422
	f3	-50.9 ± 36.1	47.4 ± 8.7	1/4	222
100	f1	-31.4 ± 8.9	$--$	0/3	8
	f2	$--$	$--$	1/4	47
	f3	-25.3 ± 9.7	35 ± 4.2	0/4	15

Double Gaussian Fit. Steps found by SFAO, 10 iterations. Histograms [-200:200] nm, bin 10nm.

Force = 0.5 pN

Table E.31.: **StepSize 100 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	N° steps
free	f1	-101.2 ± 22.4	90.3 ± 3.3	1/1	254
	f2	-88.2 ± 6.8	88.7 ± 6.0	3/3	1011
	f3	-97.4 ± 22.2	85.2 ± 36.6	3/3	982
30	f1	-110.0 ± 2.6	100.2 ± 3.2	1/1	1150
	f2	-101.9 ± 7.5	90.9 ± 29.2	3/3	3329
	f3	-94.6 ± 25.9	66.3 ± 5.2	3/3	3022
50	f1	-90.0 ± 3.4	101.4 ± 40.1	1/1	364
	f2	-107.9 ± 32.3	93.8 ± 4.8	3/3	1565
	f3	-97.2 ± 5.2	72.1 ± 7.3	3/3	948
100	f1	-100.0 ± 2.4	90.0 ± 3.1	1/1	108
	f2	-119.5 ± 15.2	107.7 ± 4.2	3/3	353
	f3	-107.0 ± 20.9	89.6 ± 31.0	3/3	345

Table E.32.: **StepSize 50 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	N° steps
free	f1	-92.8 ± 34.8	78.0 ± 41.9	3/3	691
	f2	-74.1 ± 15.4	46.6 ± 4.7	3/3	992
	f3	-82.1 ± 37.1	87.5 ± 4.3	3/3	954
30	f1	-87.4 ± 4.8	85.7 ± 34.6	3/3	2572
	f2	-87.7 ± 26.6	73.6 ± 34.1	3/3	2856
	f3	-90.3 ± 32.1	84.7 ± 5.7	3/3	2908
50	f1	-101.6 ± 33.3	86.2 ± 32.6	3/3	1348
	f2	-79.2 ± 23.5	100.0 ± 3.4	3/3	796
	f3	-81.7 ± 40.1	65.6 ± 40.1	3/3	951
100	f1	-67.3 ± 7.6	64.5 ± 8.7	1/3	129
	f2	-50.0 ± 3.2	57.3 ± 4.6	1/3	31
	f3	-50.0 ± 6.8	61.2 ± 8.3	1/3	39

Table E.33.: **StepSize 30 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	N° steps
free	f1	-92.4 ± 41.4	83.0 ± 4.6	3/3	838
	f2	-68.8 ± 10.2	76.8 ± 13.6	3/3	840
	f3	-73.0 ± 29.6	68.3 ± 34.3	3/3	807
30	f1	-93.6 ± 15.6	88.3 ± 3.3	3/3	3122
	f2	-96.6 ± 39.3	80.9 ± 22.9	3/3	3109
	f3	-77.7 ± 7.0	73.5 ± 33.0	3/3	2796
50	f1	-86.3 ± 9.9	103.8 ± 83.3	3/3	1269
	f2	-76.7 ± 4.2	100.0 ± 3.2	3/3	1186
	f3	-73.2 ± 29.0	61.7 ± 27.9	3/3	677
100	f1	-40.0 ± 3.2	40.0 ± 3.4	1/3	14
	f2	$--$	$--$	1/3	37
	f3	-26.5 ± 4.5	20.0 ± 3.2	0/3	18

Double Gaussian Fit. Steps found by SFAO, 10 iterations. Histograms [-200:200] nm, bin 10nm.

Force = 0.3 pN

Table E.34.: **StepSize 100 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	Nº steps
free	f1	-124.4 ± 4.7	150 ± 3.1	3/3	1673
	f2	-145.3 ± 5.0	121.8 ± 35.0	3/3	1713
	f3	-113.9 ± 29.2	107.5 ± 40.8	3/3	1568
30	f1	-128.3 ± 31.0	128.1 ± 40.2	3/3	3680
	f2	-115.6 ± 5.8	115 ± 5.4	3/3	3724
	f3	-96.4 ± 18.7	103.6 ± 4.9	3/3	3639
50	f1	-134.5 ± 34.7	127.8 ± 47.9	3/3	3485
	f2	-119.4 ± 18.2	114.6 ± 5.5	3/3	3223
	f3	-106.8 ± 8.1	94.89 ± 11.5	3/3	2699
100	f1	-136.2 ± 38.9	135.9 ± 4.9	3/3	721
	f2	-108.3 ± 15.7	123.9 ± 14.2	3/3	547
	f3	-103.4 ± 4.9	109.1 ± 32.3	3/3	463

Table E.35.: **StepSize 50 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	Nº steps
free	f1	-121.1 ± 12.0	133.8 ± 33.9	3/3	1302
	f2	-142.9 ± 4.5	104.9 ± 20.1	3/3	1202
	f3	-128.8 ± 3.7	104.8 ± 33.2	3/3	1745
30	f1	-124.5 ± 39.3	110.3 ± 3.9	3/3	3082
	f2	-131.1 ± 29.6	107.8 ± 23.3	3/3	3019
	f3	-110.3 ± 20.0	100.7 ± 8.3	3/3	3858
50	f1	-110.6 ± 15.0	128.7 ± 3.4	3/3	2791
	f2	-123.3 ± 50.2	98.8 ± 6.9	3/3	2636
	f3	-124.3 ± 5.4	94.0 ± 10.0	3/3	2763
100	f1	-94.0 ± 7.2	118.4 ± 47.3	3/3	463
	f2	-105.5 ± 9.2	96.0 ± 51.5	3/3	333
	f3	-103.9 ± 47.1	89.0 ± 35.6	3/3	273

Table E.36.: **StepSize 30 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	Nº steps
free	f1	-137.6 ± 14.2	114.4 ± 4.9	3/3	1502
	f2	-125.1 ± 5.0	114.3 ± 31.2	3/3	1448
	f3	-114.8 ± 11.1	104.9 ± 33.7	3/3	1529
30	f1	-114.3 ± 10.9	112.3 ± 8.7	3/3	3395
	f2	-103.6 ± 12.5	114.3 ± 39.5	3/3	3472
	f3	-109.0 ± 26.9	101.1 ± 14.8	3/3	3678
50	f1	-128.2 ± 15.8	116.5 ± 5.0	3/3	3166
	f2	-127.8 ± 29.7	115.2 ± 40.6	3/3	2711
	f3	-118.1 ± 23.5	98.6 ± 16.8	3/3	2606
100	f1	-126.6 ± 50.3	125.6 ± 51.0	3/3	395
	f2	-123.4 ± 60.8	103.3 ± 25.0	3/3	259
	f3	-101.8 ± 47.1	93.1 ± 20.3	3/3	170

Double Gaussian Fit. Steps found by SFAO, 10 iterations. Histograms [-200:200] nm, bin 10nm.

SMOOTH 10 Force = 1 pN

Table E.37.: **StepSize 100 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	N° steps
free	f1	-41.3 ± 20.2	30.4 ± 15.4	3/3	1655
	f2	-38.9 ± 27.9	20.7 ± 9.4	4/4	3433
	f3	-35.5 ± 14.1	27.5 ± 13.3	4/4	3185
50	f1	-100.6 ± 7.0	93.0pm3.9	3/3	359
	f2	-92.9 ± 6.5	83.7 ± 6.2	4/4	756
	f3	-87.0 ± 6.8	78.0 ± 18.2	4/4	838
100	f1	-100.0 ± 2.8	93.3 ± 4.1	3/3	362
	f2	-94.2 ± 9.6	83.6 ± 4.8	4/4	537
	f3	-89.1 ± 11.5	78.5 ± 9.9	4/4	551

Table E.38.: **StepSize 50 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	N° steps
50	f1	-33.1 ± 14.4	21.6pm9.0	3/3	1654
	f2	-22.7 ± 5.4	18.2 ± 6.4	4/4	3823
	f3	-31.9 ± 20.7	19.5 ± 11.3	4/4	2952
50	f1	-53.5 ± 7.2	46.3 ± 8.1	3/3	437
	f2	-53.0 ± 5.5	37.2 ± 4.2	4/4	606
	f3	-49.7 ± 14.4	39.1 ± 10.6	4/4	646
100	f1	-54.5 ± 4.5	45.0 ± 7.9	3/3	370
	f2	-56.4 ± 10.45	47.1 ± 4.2	4/4	363
	f3	-45.1 ± 6.3	37.3 ± 7.3	4/4	299

Table E.39.: **StepSize 30 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	N° steps
free	f1	-29.3 ± 13.0	18.8 ± 9.6	3/3	1749
	f2	-25.5 ± 10.8	13.7 ± 6.8	4/4	3350
	f3	-23.5 ± 7.8	21.5 ± 10.9	4/4	2708
50	f1	-23.4 ± 5.5	17.2 ± 5.2	2/3	56
	f2	-35.0 ± 26.7	40.3 ± 37.7	3/4	267
	f3	-38.7 ± 7.1	25.9 ± 4.9	3/4	152
100	f1	-30.0 ± 6.7	13.8 ± 4.7	1/3	15
	f2	——	——	1/4	35
	f3	-24.8 ± 13.1	30.0 ± 11.6	0/4	24

Double Gaussian Fit. Steps found by SFAO, 10 iterations. Histograms [-200:200] nm, bin 10nm.

SMOOTH 10 Force = 0.5 pN

Table E.40.: **StepSize 100 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	N° steps
free	f1	-63.98 ± 4.8	51.0 ± 19.9	1/1	808
	f2	-58.4 ± 10.5	52.4 ± 20.2	3/3	2739
	f3	-53.1 ± 16.9	44.0 ± 19.5	3/3	2649
50	f1	-110.0 ± 3.3	84.0 ± 9.8	1/1	151
	f2	-92.4 ± 46.2	83.3 ± 28.5	3/3	560
	f3	-95.0 ± 4.7	62.4 ± 6.1	3/3	481
100	f1	-110.0 ± 3.0	100.0 ± 2.3	1/1	107
	f2	-117.8 ± 10.7	103.1 ± 6.7	3/3	347
	f3	-90.1 ± 24.1	78.8 ± 19.2	3/3	355

Table E.41.: **StepSize 50 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	N° steps
free	f1	-62.2 ± 22.1	52.1 ± 17.2	3/3	2076
	f2	-47.5 ± 4.2	37.8 ± 17.0	3/3	2598
	f3	-62.0 ± 5.9	37.7 ± 9.7	3/3	2246
50	f1	-73.9 ± 31.7	64.7 ± 14.4	3/3	583
	f2	-53.9 ± 8.5	51.5 ± 19.2	3/3	390
	f3	-56.5 ± 7.1	42.7 ± 8.0	3/3	445
100	f1	-86.8 ± 4.5	70.0 ± 3.2	3/3	71
	f2	-47.2 ± 4.3	45.5 ± 4.5	2/3	34
	f3	-46.9 ± 5.7	60.0 ± 3.3	1/3	48

Table E.42.: **StepSize 30 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	N° steps
free	f1	-59.2 ± 9.1	42.1 ± 9.2	3/3	2249
	f2	-56.8 ± 8.9	47.8 ± 6.5	3/3	2263
	f3	-51.3 ± 6.1	37.8 ± 8.3	3/3	1998
50	f1	-55.8 ± 24.8	44.3 ± 9.2	3/3	447
	f2	-59.3 ± 31.2	52.3 ± 33.1	3/3	425
	f3	-46.7 ± 5.4	46.7 ± 6.1	3/3	259
100	f1	-21.5 ± 66.2	50.1 ± 0.6	1/3	24
	f2	-60.0 ± 3.2	50.0 ± 3.1	2/3	57
	f3	-34.5 ± 3.8	56.3 ± 4.9	2/3	24

Double Gaussian Fit. Steps found by SFAO, 10 iterations. Histograms [-200:200] nm, bin 10nm.

SMOOTH 10 Force = 0.3 pN

Table E.43.: **StepSize 100 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	N° steps
free	f1	-87.4 ± 28.1	79.4 ± 14.7	3/3	2784
	f2	-82.3 ± 22.1	60.5 ± 15.5	3/3	2789
	f3	-81.5 ± 16.8	72.5 ± 24.6	3/3	2797
50	f1	-100.4 ± 12.9	83.1 ± 15.8	3/3	1921
	f2	-92.4 ± 31.3	85.5 ± 31.3	3/3	1614
	f3	-96.2 ± 17.3	80.8 ± 34.9	3/3	1374
100	f1	-122.7 ± 4.1	115.5 ± 5.0	3/3	483
	f2	-120.0 ± 4.7	120.0 ± 3.0	3/3	429
	f3	-95.4 ± 4.9	90.0 ± 3.1	3/3	374

Table E.44.: **StepSize 50 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	N° steps
free	f1	-78.7 ± 3.5	73.6 ± 21.2	3/3	2532
	f2	-83.3 ± 6.9	68.8 ± 25.9	3/3	2658
	f3	-78.3 ± 19.1	59.9 ± 14.6	3/3	2702
50	f1	-78.9 ± 3.6	89.8 ± 15.0	3/3	1606
	f2	-86.6 ± 6.8	78.2 ± 6.8	3/3	1436
	f3	-89.7 ± 21.3	80.6 ± 24.8	3/3	1284
100	f1	-97.5 ± 42.8	96.6 ± 4.6	3/3	262
	f2	-105.4 ± 4.9	60.0 ± 2.8	3/3	157
	f3	-74.1 ± 4.9	70.5 ± 7.0	3/3	181

Table E.45.: **StepSize 30 nm**

Noise Code	Focal distance	Negative peak (nm)	Positive Peak (nm)	Traces LLR ok	N° steps
free	f1	-86.9 ± 22.5	64.1 ± 7.6	3/3	2624
	f2	-80.3 ± 11.4	69.7 ± 24.9	3/3	2573
	f3	-77.7 ± 9.0	63.3pm7.9	3/3	2458
50	f1	-80.0 ± 3.4	80.0 ± 2.8	3/3	1666
	f2	-88.6 ± 21.9	95.1 ± 10.4	3/3	1368
	f3	-100.0 ± 3.4	77.3 ± 29.0	3/3	1241
100	f1	-86.3 ± 5.1	75.0 ± 4.6	3/3	165
	f2	-90.6 ± 11.0	93.1 ± 4.5	3/3	142
	f3	-98.5 ± 26.6	76.7 ± 4.5	3/3	115

Double Gaussian Fit. Steps found by SFAO, 10 iterations. Histograms [-200:200] nm, bin 10nm.

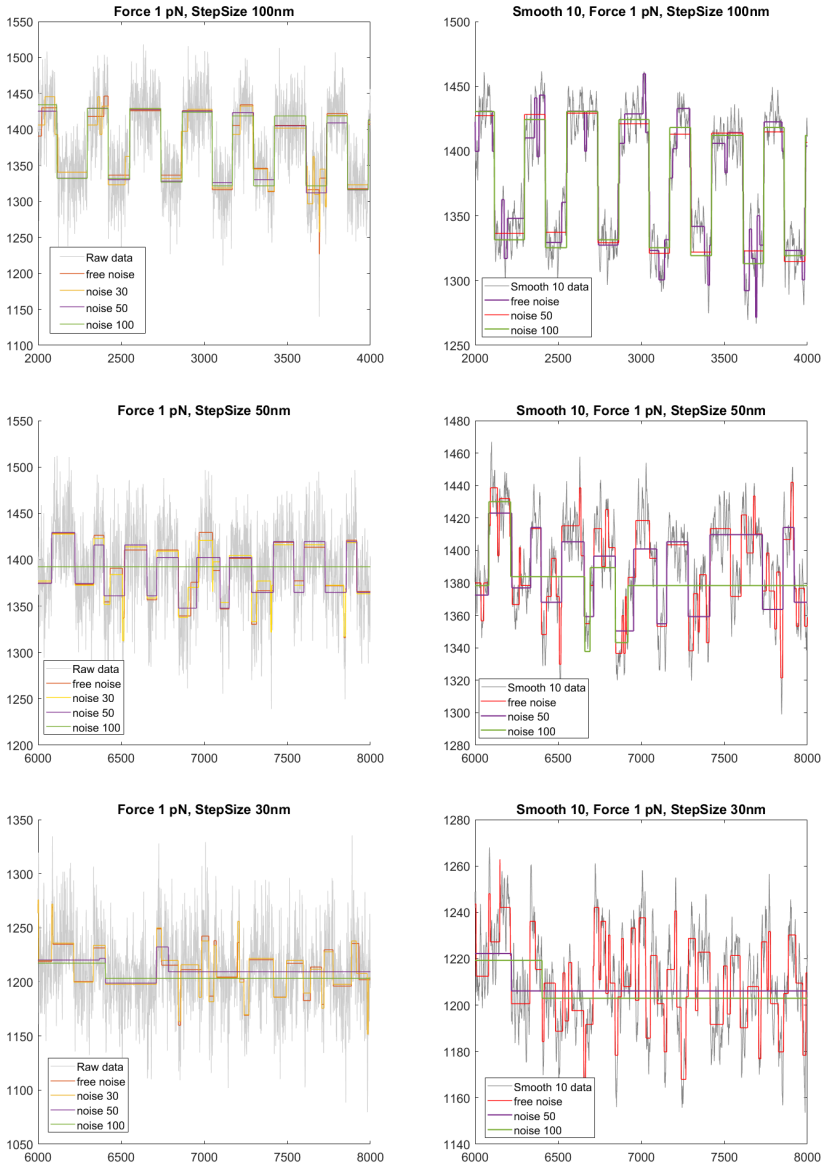

Figure E.13.: Representative trace of a Control DNA in *Up and Down experiment* , Force 1 pN. Steps found by SFAO for Raw data (Right) with a measnoise = [free, 30, 50, 100] and for Smooth by 10, gaussian data (Left). with a measnoise = [free, 50, 100]. x-bar legth ~ 17s

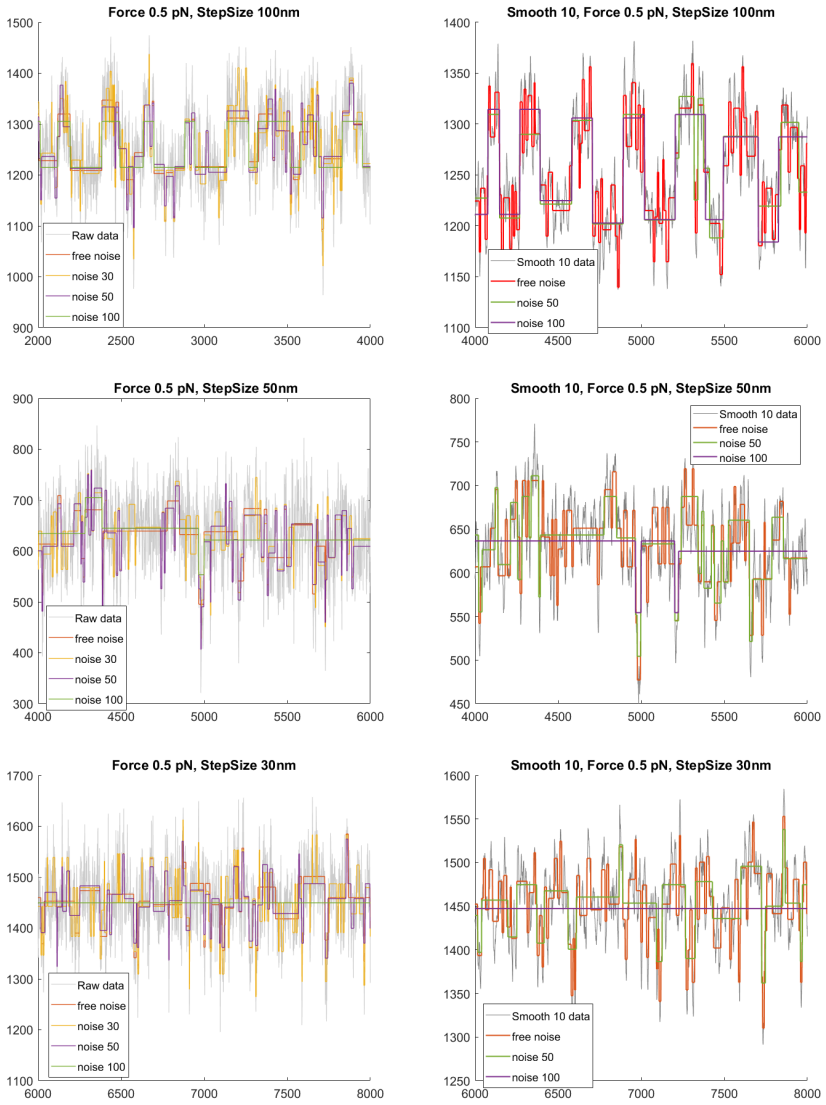

Figure E.14.: Representative trace of a Control DNA in *Up and Down experiment* , Force 0.5 pN. Steps found by SFAO for Raw data (Right) with a measnoise = [free, 30, 50, 100] and for Smooth by 10, gaussian data (Left). with a measnoise = [free, 50, 100]. x-bar legth ~ 17s

E.5 MIXED STEPS:SFA VS SFAO

The MT traces could be formed by steps of different height, for this reason a artificial step-like trace has been acquired with steps from 100 to 20 nm, with different time length:

- Total DNA tracked: 6.
- Standard Flow Cells, 1 µm beads, DNA G12E, 1mm gap magnets.
- Forces considered: 0.5 pN.
- Stepsize considered: 100, 50, 30, 20 and 10 nm.
- Algorithm employed: SFAO (measnoise 100 and 200) and SFA (2-iter).

E.5.1 *SFA*

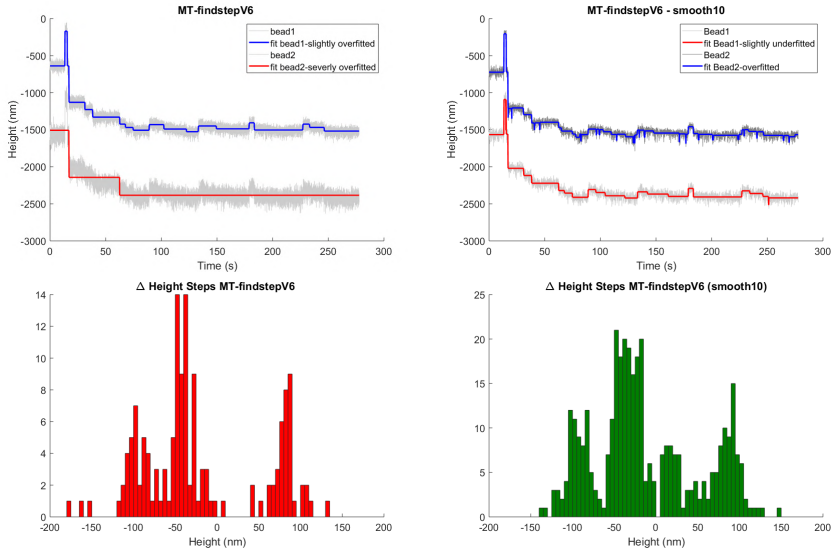

Figure E.15.: Top. Two example of data fitted by SFA for raw data (right) and data smooth by 10 (left). Bottom. The Δ height histograms for the 10 curves.

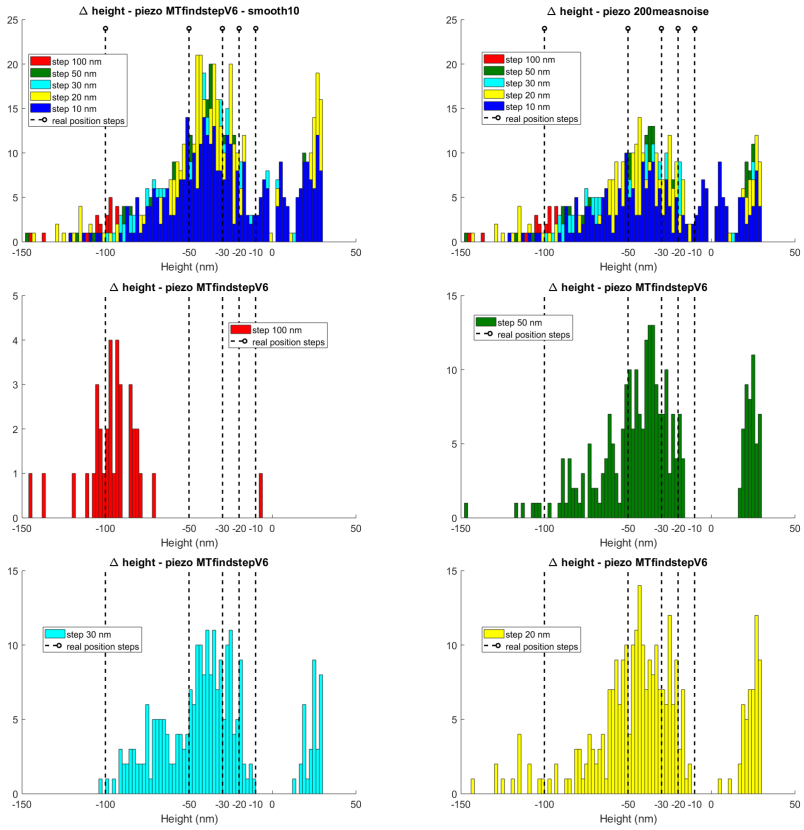

Figure E.16.: Top. The Δ height histograms for the 10 curves processed by the *SFA* for 2-iterations cycle (Right) and 1-iteration cycle (Left). Data are smooth by 10. Bottom. Single histograms taken from the Top-Right graph.

E.5.2 *SFAO*

measnoise: 100 vs 200

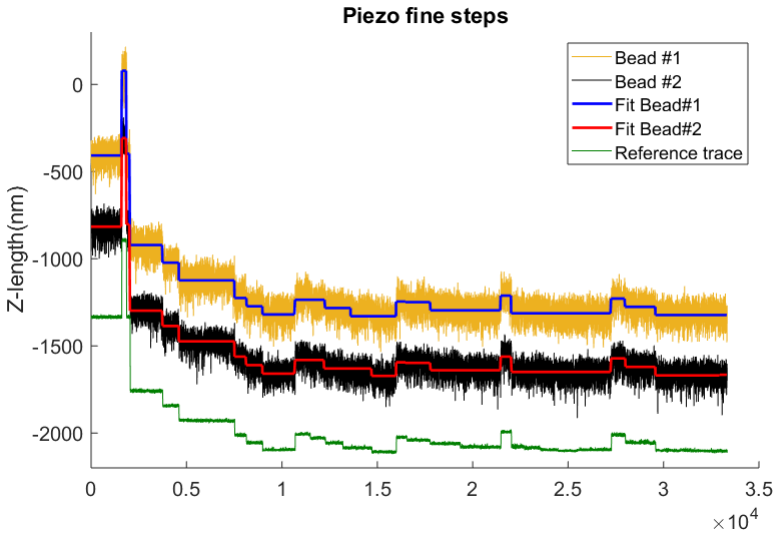

Figure E.17.: Two beads with their step-fit. In green the reference trace.

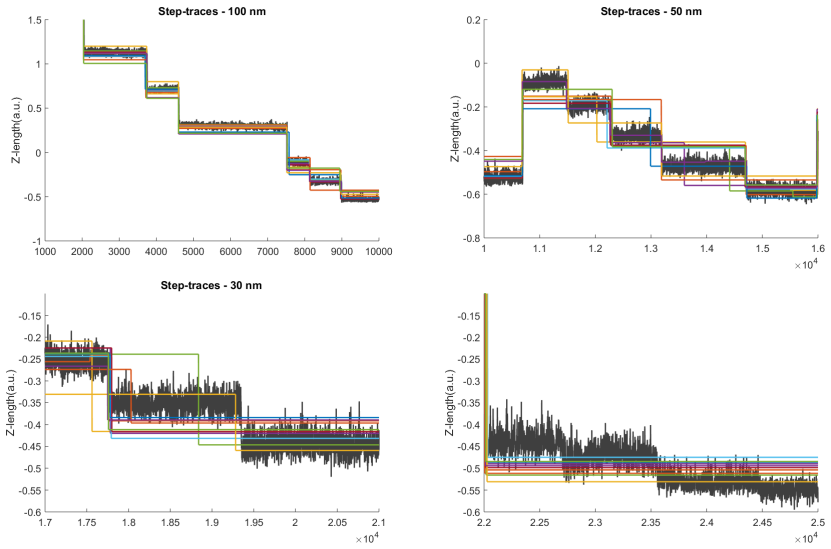

Figure E.18.: Ten step-fit for the ten traces for the different step-size. Data obtained divided the curves to have 3/4 consecutive steps of the same height.

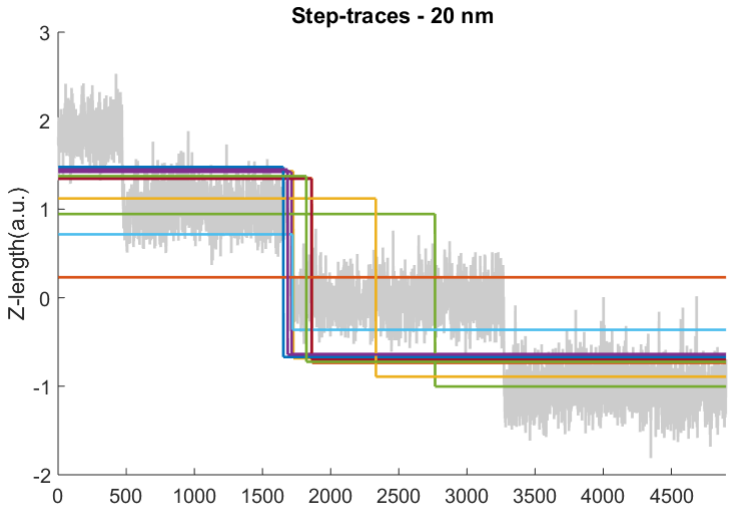

Figure E.19.: The 20-nm-steps-fits for the traces (in gray the reference bead). The algorithm detected less steps than expected.

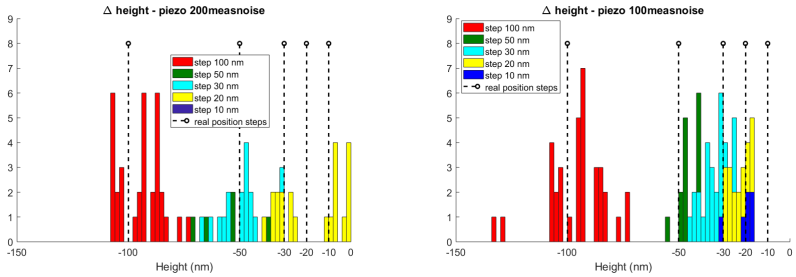

Figure E.20.: Histograms output *meansnoise* 200 (left) and *meansnoise* 100 (right) vs the real input steps (black).[10 curves, ~ 50 nm sigma noise, F = 0.5pN]

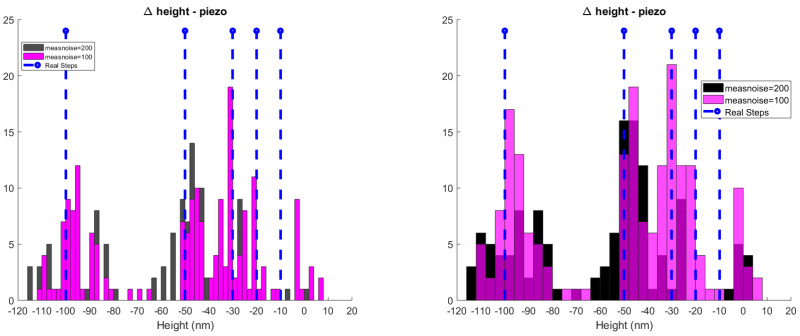

Figure E.21.: Histograms output *meansnoise* 200 (black) and *meansnoise* 100 (pink) vs the real input steps (blue). Histos bin 2 nm (Left) and 4 nm (Right). [10 curves, ~ 50 nm sigma noise, F = 0.5pN]

	Real Input Steps				
	100 nm	50 nm	30 nm	20 nm	10 nm
SFAO°	98.94 ± 18.10	43.10 ± 5.12	31.81 ± 6.42	22.37 ± 4.38	20.54 ± 5.08
SFAO*	94.11 ± 9.34	51.52 ± 10.53	44.01 ± 13.86	17.78 ± 14.37	NaN
SFA	76.6 ± 64.3	-2.6 ± 51.1	-3.4 ± 52.5	-1.3 ± 52.9	-0.3 ± 53.8

Table E.46.: The ΔHeight of the output steps vs the *real* input values for a 3/4 consecutive steps of the same height found by means of both SFA (° *measnoise* 100,* *measnoise* 200) and SFAO (smooth 10, 2-iterations) algorithms. For the **noise 200** no steps of 10-nm-height have been detected.

	# Input Steps detected				
	100 nm	50 nm	30 nm	20 nm	10 nm
	3	2	4	3	3
SFAO *100*	3	2	3.1	2.1	0.5
SFAO *200*	3	1.2	2.1	1.9	0
SFA *smooth10*	3.4	35.9	34.2	41.9	33.4

Table E.47.: The number of the output steps detected by the SFAO and SFA algorithm. For the **noise 200** no steps of 10-nm-height have been detected.

The extensive analysis of the previously reported instrumental conditions led us to choose the SFA algorithm as the most reliable for obtaining the step size of condensin mediate DNA collapsing. The final control experiment was performed as usual, by manually adding the *fake* artificial step to the data, by moving the piezo up and down.

- Total DNA tracked: 11
- Standard Flow Cells, 1 μm beads, DNA G12E, 1mm gap magnets.
- 3 different focal positions: f1 (~ 5μm from fully focused bead),f2 (~ 6μm)and f3 (~ 7μm).
- Forces considered: 1pN and 0.5 pN.
- Stepsize considered: 100, 50 and 30 nm.
- Algorithm employed: SFA, 2-iteration process.
- Smoothing: 10-points-gaussian-filter.

Final Control

Table E.48.: **Force 1 pN** - SFA - Smooth 10, 2-iterations cycle

StepSize (nm)	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
100	-79.1 ± 54.2	77.9 ± 28.1	11	2072(188)
50	-46.7 ± 20.7	36.7 ± 19.8	11	2143(195)
30	-32.9 ± 13.3	22.4 ± 12.8	11	2089(190)

Table E.49.: **Force 0.5 pN** - SFA - Smooth 10, 2-iterations cycle

StepSize (nm)	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
100	-72.2 ± 41.0	62.7 ± 39.0	7	1340(191)
50	-61.9 ± 28.9	50.6 ± 27.6	9	1474(164)
30	-55.62 ± 29.8	44.0 ± 27.4	9	1154(128)

Table E.50.: Values of the steps found by means of SFA (smooth 10, 2-iter) by grouping all the three different focal positions.

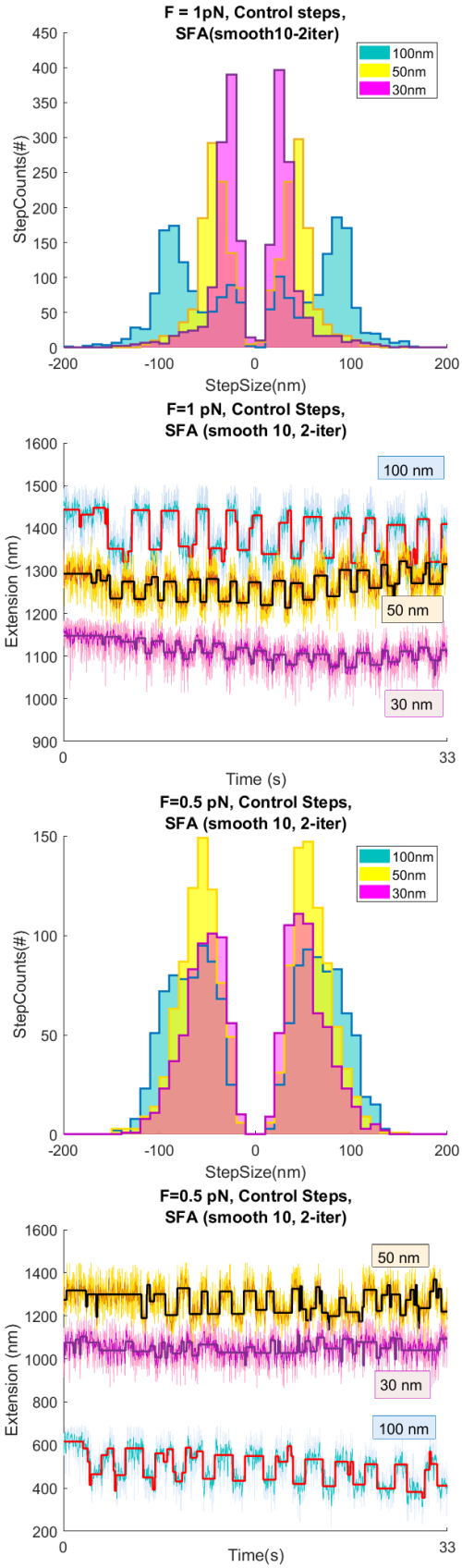

Figure E.22.: Step-Size distributions of piezo-controlled steps (up and Down) calculated by means of *SFA* (smooth10, 2-iteration cycle). StepSize 100 (Light Blue), 50 (Orange) and 30(pink)nm. Top: Force = 1pN, Bottom: Force = 0.5pN

Final Control and Focal Position contribution

Force 1 pN - SFA Code - Smooth 10, 2-iterations cycle

StepSize (nm)	Focal Distance	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
100	f1	-101.7 ± 11.5	89.1 ± 10.2	3	538(179)
	f2	-89.2 ± 24.9	77.7 ± 23.7	4	756(189)
	f3	-84.6 ± 21.8	70.4 ± 30.0	4	778(195)
StepSize (nm)	Focal Distance	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
50	f1	-51.2 ± 18.6	40.2 ± 20.3	3	585(195)
	f2	-48.3 ± 19.3	37.7 ± 18.8	4	786(197)
	f3	-42.2 ± 17.8	33.4 ± 16.2	4	772(193)
StepSize (nm)	Focal Distance	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
30	f1	-32.8 ± 12.7	22.7 ± 12.2	3	581(194)
	f2	-33.1 ± 14.2	22.7 ± 14.0	4	755(189)
	f3	-33.0 ± 13.3	22.40 ± 12.0	4	753(188)

Force 0.5 pN - SFA - Smooth 10, 2-iterations cycle

StepSize (nm)	Focal Distance	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
100	f1	-52.1 ± 24.4	43.9 ± 22.2	1	193(193)
	f2	-70.3 ± 39.7	61.1 ± 37.2	3	578(193)
	f3	-82.7 ± 37.6	72.8 ± 37.5	3	569(190)
StepSize (nm)	Focal Distance	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
50	f1	-65.7 ± 31.25	57.8 ± 31.3	3	569(190)
	f2	-60.7 ± 25.7	48.0 ± 24.4	3	351(117)
	f3	-59.2 ± 29.4	46.3 ± 24.4	3	554(185)
StepSize (nm)	Focal Distance	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
30	f1	-58.2 ± 28.8	48.6 ± 25.9	3	575(192)
	f2	-60.4 ± 32.5	46.1 ± 31.1	3	567(189)
	f3	-52.9 ± 27.0	42.9 ± 25.4	3	587(196)

Table E.51.: Values of the steps found by means of SFA (smooth 10, 2-iter) dividing the different focal contributions.

Finally, a routine experiment has been performed at high force (F = 4 pN). which is barely near the maximum force which can be exercised on a 2µm long DNA with and commercial 1µm paramagnetic beads in the setup (see Sec. 13.4). The final control experiment was performed as usual, by manually adding the *fake* artificial step to the data, by moving the piezo up and down.

- Total DNA tracked: 15
- Standard Flow Cells, 1 µm beads, DNA G12E, 1mm gap magnets.
- Forces considered:4 pN.
- Stepsize considered: 100, 50 and 30 nm.
- Algorithm employed: SFA, 2-iteration process.
- Smoothing: 10-points-gaussian-filter.

Final Control: maximum precision

Table E.52.: **Force 4 pN** - SFA - Smooth 10, 2-iterations cycle

StepSize (nm)	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
100	-48.8 ± 28.6	37.8 ± 22.4	15	2330(155)
50	-32.06 ± 17.0	22.3 ± 14.1	15	1426(95)
30	-24.3 ± 11.3	17.2 ± 17.3	15	1423(95)

Table E.53.: **Force 4 pN** - SFA - NO Smooth, 2-iterations cycle

StepSize (nm)	Negative peak (nm)	Positive Peak (nm)	Traces	N steps
100	-92.1 ± 56.7	78.5 ± 54.5	15	1354(90)
50	-47.1 ± 25.3	39.4 ± 26.3	15	1424(95)
30	-35.9 ± 18.4	27.2 ± 21.7	15	1547(103)

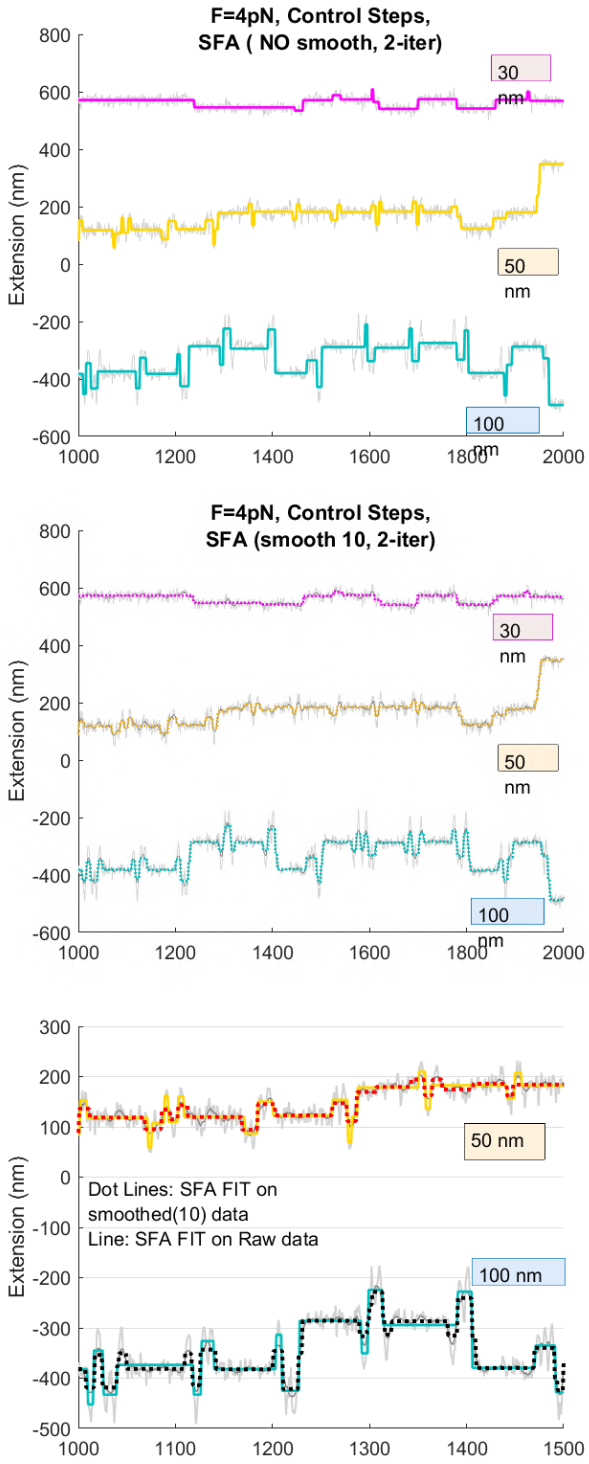

Figure E.23.: Representative Traces for fast-piezo-controlled experiments steps (up and Down calculated by means of *SFA* 2-iteration cycle). StepSize 100 (Light Blue), 50 (Orange) and 30(pink)nm. Top: No smooth, Middle: 10-pts gaussian smooth, Bottom: no smooth vs 10-pnts-smooth.

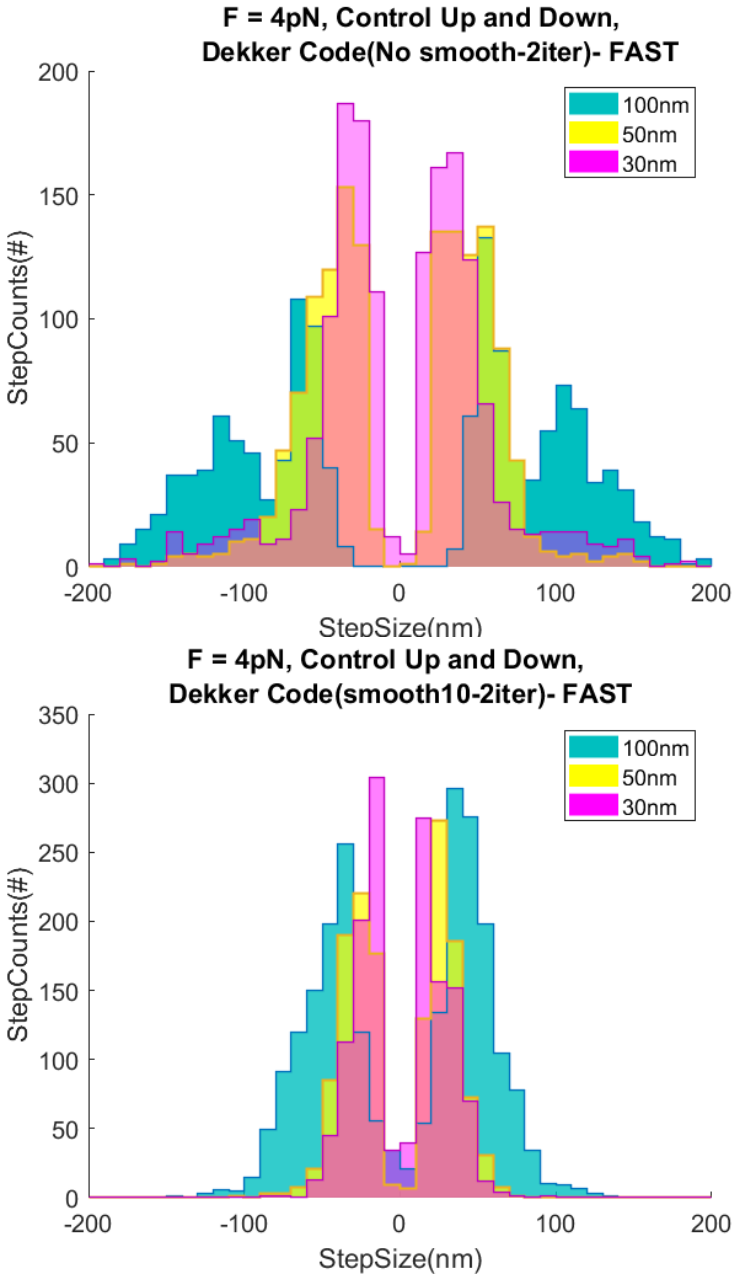

Figure E.24.: Step-Size distributions of piezo-controlled steps (up and Down) calculated by means of *SFA* (2-iteration cycle). Step-Size 100 (Light Blue), 50 (Orange) and 30(pink)nm. Top: No smooth, Bottom: 10-pts gaussian smooth.

The SFAO algorithm had been tested also on different protein:mediated collapsed, in particular parB protein [43, 294]. Data were obtained by Dr. Cesar Lopez-Pastrana and partially published in [294].

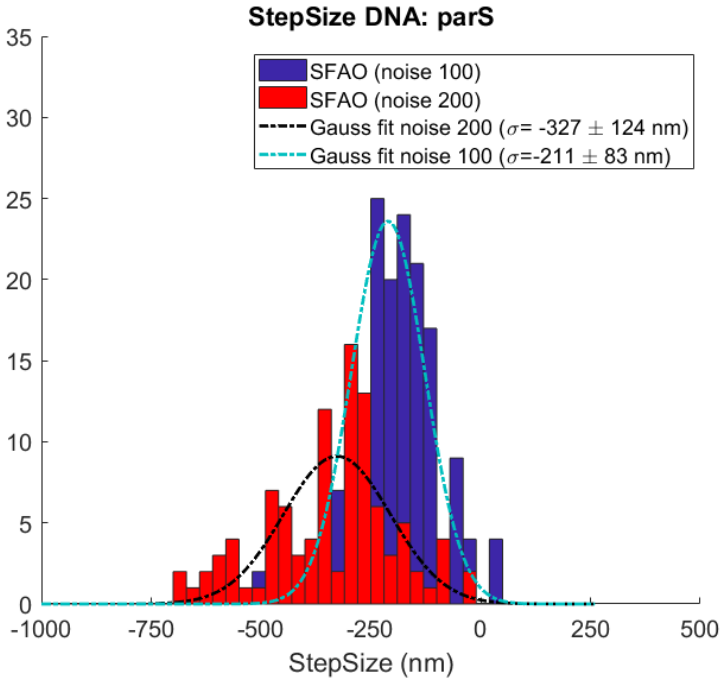

Figure E.25.: ΔHeight steps for DNA condensation by ParB obtained by means of the SFAO (Blue, Measnoise=100 and Red, Measnoise=200). In the legend are reported the means±std obtained for a single (parS) Gaussian fit (StepSize=50nm).

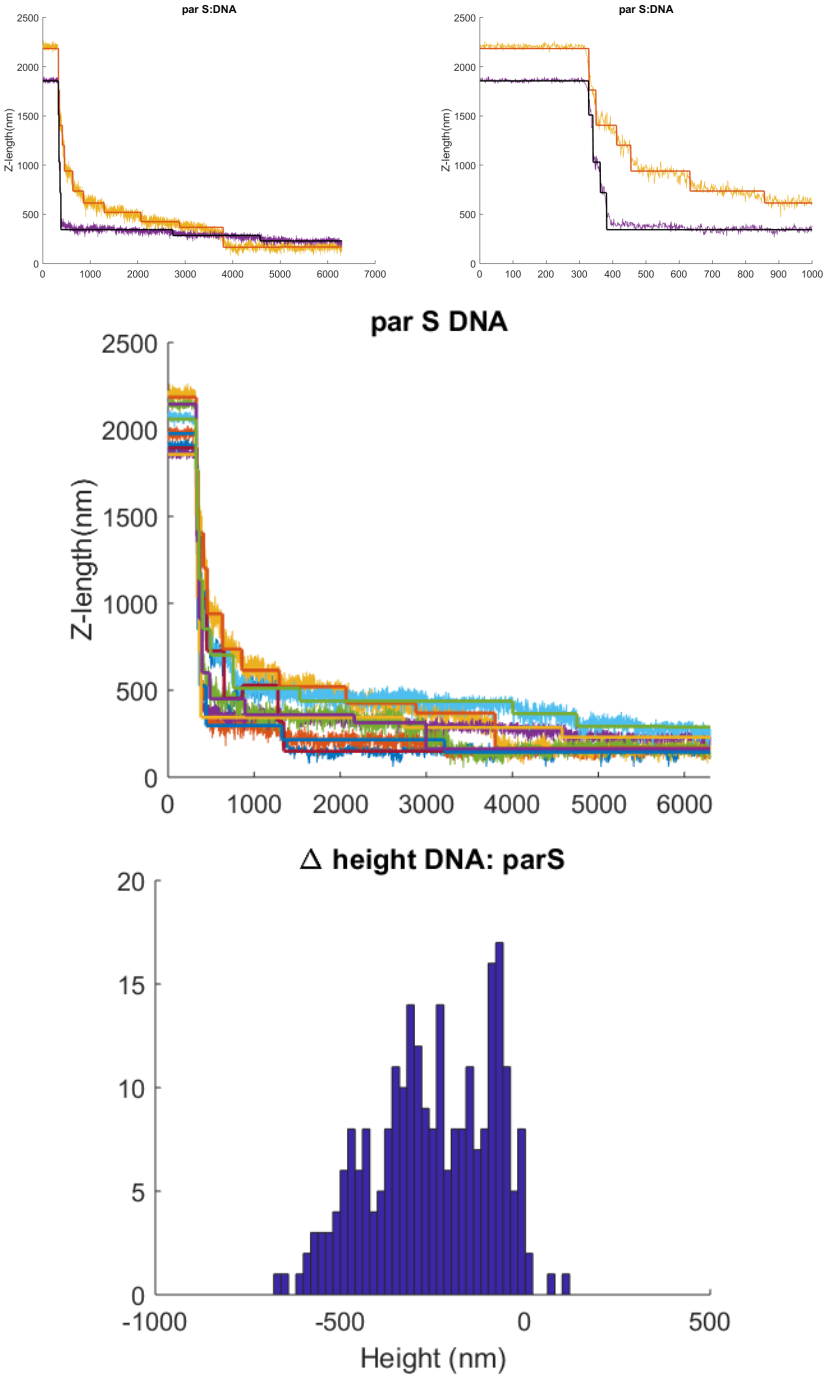

Figure E.26.: Representative traces of parB protein collapsing DNA (pET28:parS) and StepSize found by means of SFAO (N traces = 36, binning = 20 nm, LLR values [0.2506:13.3564]).

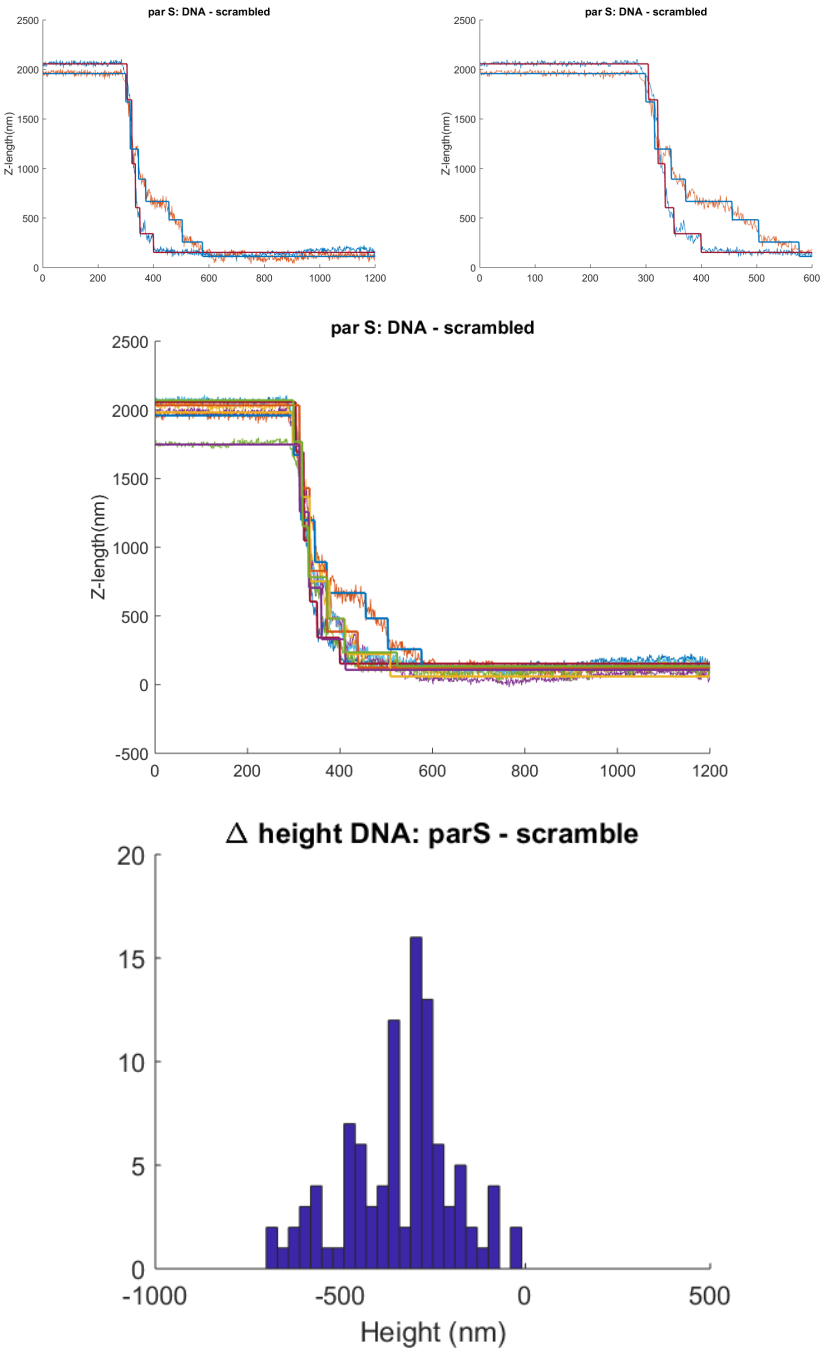

Figure E.27.: Representative traces of parB protein collapsing DNA (scrambled:parS) and StepSize found by means of SFAO (N traces = 36, binning = 20 nm, LLR values [0.2506 : 13.3564]).

F

λ/2 DNA:CONDENSIN INTERACTIONS

F.1 CONDENSIN 30NM - 8.6μM-LONG λ/2 DNA

The two SFA and SFAO software were applied also to data of λ/2 DNA condensation in the presence of 30 nM of condensin. All the data showed here were obtained by Dr. Silvia Hormeño-Torres with a protocol very similar to the one described in Chapter14 (with condensin 30 nM, ATP 1mM, force F=0.75 pN, fitted traces in Fig.F.2). With respect to the data presented in the main part, these results could be directly compared to the one previously published [254] because the λ/2 DNA has a length (~ 8.6μm) very similar to the one employed by Dekker and colleagues (~ 6.5μm).

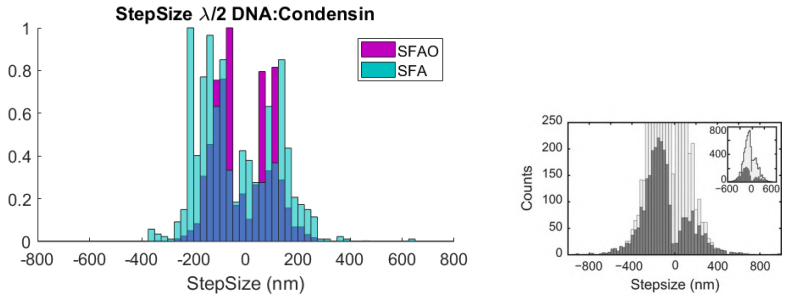

Figure F.1.: ΔHeight steps for DNA condensation by Condensin obtained by means of the SFA (light blue) and by SFAO (purple). For comparison purpose, the step-size histogram published by Dekker in [254] is reported.

Table F.1.: **Force 0.75 pN**

Algorithm	*measnoise*(nm)	Negative peak(nm)	Positive Peak(nm)
SFAO	100	-104 ± 42	78 ± 43
SFAO	200	-153 ± 69	104 ± 68
SFA	2-iter, smooth10	-114.6 ± 70.5	94.85 ± 70.96

Table F.2.: StepSize for a double gaussian fit for λ/2DNA:condensin interactions for SFAO (*measnoise* 100 and *measnoise* 200) and SFA(2-iter, smooth-10).

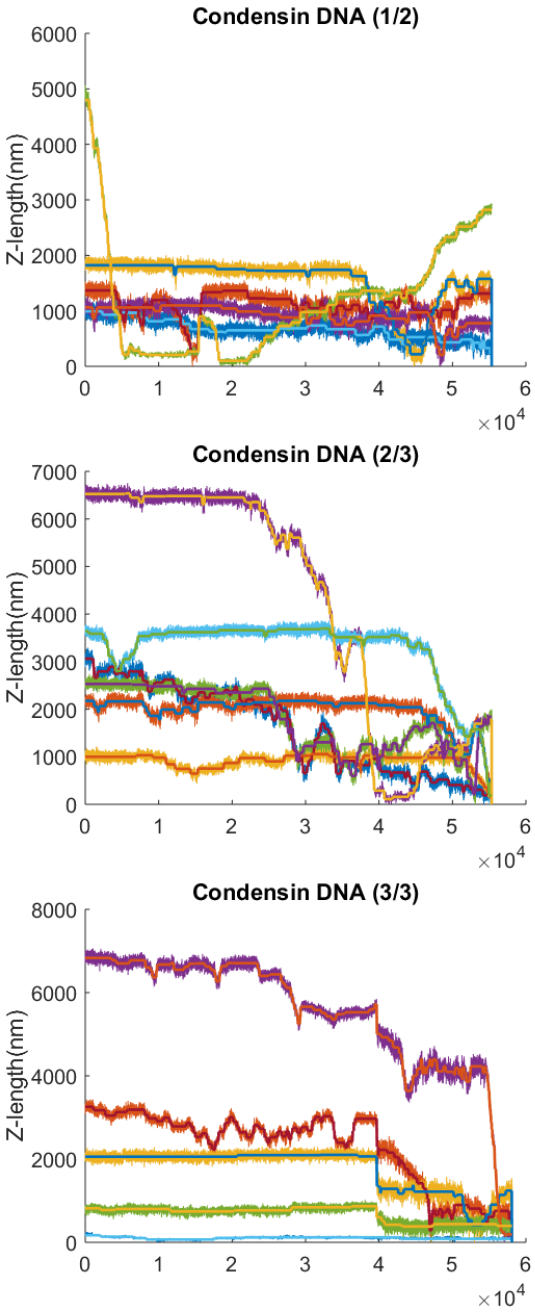

Figure F.2.: Representative curve of DNA:condensin interaction and step-fit by SFAO (measnoise 200). DNA λ/2, 30nM condensin, F=0.75pN.

G

FLOW STRETCHING: ADDITIONAL DATA

G.1 BOUNDARY EXCHANGE: CONSTRAINED AND FREE FIT

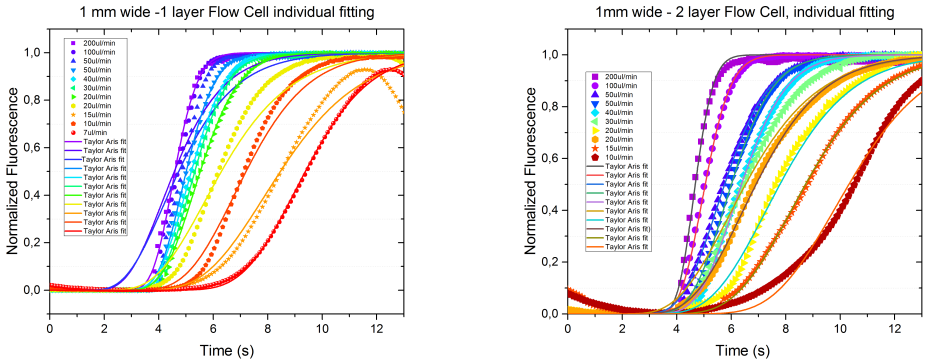

Figure G.1.: Normalized fluorescence vs time at different flow rates Q for (Left) 1 mm wide - 1 layer and (Right) 1 mm wide - 2 layer. Data were independently fitted to Taylor-Aris formula.

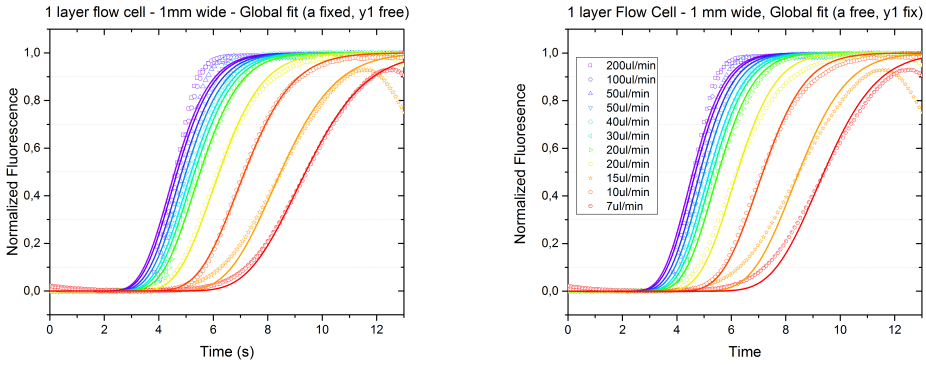

Figure G.2.: Normalized fluorescence vs time at different flow rates Q for 1 mm wide - 1 layer. Data were globally fitted to Taylor-Aris formula. For both the figures: ν free, $D = 4.9 \cdot 10^{-10}$. (Left) $a = 100\mu m$ parameter fixed and $y_1 = 0.095mm$ obtained from the global fit and (Right) $y_1 = 1mm$ parameter fixed and $a = 830mm$ obtained from the fit.

Flow Rate ($\mu l/min$)	y_0 (m)	v (m/s)	D (m^2/s)	a (m)
200	8.10E-04	1.73E-04	8.61E-10	8.63E-05
100	8.30E-04	1.64E-04	1.20E-09	4.76E-05
50	8.27E-04	1.44E-04	1.44E-09	1.46E-04
50	8.26E-04	1.37E-04	2.10E-09	5.24E-05
40	7.92E-04	1.24E-04	3.68E-10	9.08E-05
30	8.65E-04	1.33E-04	2.87E-09	3.15E-04
20	7.42E-04	9.53E-05	1.22E-09	1.86E-04
20	6.80E-05	1.26E-04	2.10E-08	2.42E-04
15	0.00319	3.66E-04	2.84E-09	3.76E-04
10	0.00105	1.02E-04	5.81E-10	1.51E-04
Mean	6.30E-04		1.54E-05	1.28E+00
Theoretical	5.00E-04		4.90E-10	1.40E-04

Table G.1.: 1 mm wide - 2 layer flow cell independent Taylor-Aris fit: 4 free parameters: y_1 (distance from inlet), v (fluid speed), a (channel height), D (diffusion constant).

Flow Rate ($\mu l/min$)	y_0 (m)	v (m/s)	D (m^2/s)	a (m)
200	7.97E-04	1.73E-04	1.43E-09	7.80E-07
100	7.91E-04	1.73E-04	3.31E-09	2.77E-04
50	8.04E-04	1.70E-04	4.01E-09	3.43E-04
50	8.22E-04	1.64E-04	1.35E-09	7.06E-05
40	1.04E-03	2.00E-04	3.63E-09	3.20E-07
30	9.10E-04	1.71E-04	1.91E-09	9.72E-05
20	8.80E-04	1.59E-04	2.50E-09	8.45E-05
20	8.30E-04	1.31E-04	2.71E-09	2.99E-04
15	9.00E-04	1.06E-04	1.75E-09	2.38E-04
10	8.67E-04	1.20E-04	1.59E-09	1.91E-04
7	9.87E-04	1.05E-04	9.95E-10	1.37E-04
Mean	8.71E-04		2.08E-09	5.81E-05
Theoretical	1.00E-03		4.90E-10	1.00E-04

Table G.2.: 1 mm wide - 1 layer flow cell independent Taylor-Aris fit: 4 free parameters: y_1 (distance from inlet), v (fluid speed), a (channel height), D (diffusion constant).

BIBLIOGRAPHY

[1] R. Corti, C. A. Marrano, D. Salerno, S. Brocca, A. Natalello, C. Santambrogio, G. Legname, F. Mantegazza, R. Grandori and V. Cassina. *Depicting confoormational ensembles of a-synuclein by single molecule force spectroscopy and native mass spectroscopy*, International Journal of Molecular Sciences, **20**, (20), 5181, (2019).

[2] M. Cristofalo, D. Kovari, R. Corti, D. Salerno, V. Cassina, D. Dunlap and F. Mantegazza. *Nanomechanics of Diaminopurine-Substituted DNA*, Biophysical Journal, **116**, 760-771, (2019).

[3] I. Figuereido, A. Paiotta, R. Dal Magro, F. Tinelli, R. Corti, F. Re, V. Cassina, E. Caneva, F. Nicotra and L. Russo. *A New Approach for Glyco-Functionalization of Collagen-Based Biomaterials*, International Journal of Molecular Sciences, **20**, 1747, (2019).

[4] R. Dal Magro, S. Simonelli, A. Cox, B. Formicola, R. Corti, V. Cassina, L. Nardo, F. Mantegazza, D. Salerno, G. Grasso, M. A. Deriu, A. Danani, L. Calabresi and F. Re. *The extent of human apolipoprotein A-I lipidation strongly affects the beta amyloid efflux across the blood-brain barrier in vitro*, Frontiers in Neuroscience, doi: 10.3389/fnins.2019.00419, (2019).

[5] B. Torsello, S. De Marco, S. Bombelli, E. Chisci, V. Cassina, R. Corti, D. Bernasconi, R. Giovannoni, C. Bianchi and R. A. Perego. *The 1ALCTL and 1BLCTL isoforms of Arg/Abl2 induce fibroblast activation and extra cellular matrix remodelling differently*, Biology Open, **8**, bio038554, (2019).

[6] V. Di Stefano, B. Torsello, C. Bianchi, I. Cifola, E. Mangano, G. Bovo, V. Cassina, S. De Marco, R. Corti, C. Meregalli, S. Bombelli, P. Viganò, C. Battaglia, G. Strada and R. A. Perego. *Major Action of Endogenous Lysyl Oxidase in Clear Cell Renal Cell Carcinoma Progression and Collagen Stiffness Revealed by Primary Cell Cultures*, The American Journal of Pathology, **9**, Vol. 186, (2016).

[7] D. Fontana, M. Mauri, R.Renso, M. Docci, I. Crespiatico, L. Røst, A. Niro, D. D'Aliberti, L. Massimino, M. Bertagna, G. Zambrotta, M. Bossi, S. Citterio, B. Crescenzi, F. Fanelli, V. Cassina, R. Corti, D. Salerno, L. Nardo, F. Mantegazza, C. Mecucci, G. Cavaletti, P. Bruheim, D. Rea, S. Larsen, C. Gambacorti-Passerini, R. Piazza. *ETNK1 mutations induce a mutator phenotype that can be reverted with phosphoethanolamine*, Nature Communication, under review, (2019).

[8] R. Corti , A. Cox , V. Cassina, L. Nardo, D. Salerno, C. A. Marrano, N. Missana, P. Andreozzi, P. J. Silva, F. Stellacci, R. Dal Magro, F. Re, F. Mantegazza. *Clustering on gold nanoparticle surface improves the anti-amyloidogenic effect of mApoE peptide*, under preparation.

[9] M. Cristofalo, C. A. Marrano, D. Salerno, R. Corti, V. Cassina, B. Sclavi, A. Mammola, M. Gherardi, M. Cosentino Lagomarsino, F. Mantegazza *Cooperative effects on the compaction of short DNA fragments by the nucleoid protein H-NS and the crowding agent PEG probed by Magnetic Tweezers*, under preparation.

[10] K. C. Neuman and A. Nagy, *Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy*, Nat. Methods, **5**(6), 491-505, (2008).

[11] T. Ha, *Single-molecule methods leap ahead*, Nat. Methods, **11**, 1015-1018, (2014).

[12] N. G. Walter, *Single Molecule Detection, Analysis, and Manipulation*, Encyclopedia of Analytical Chemistry R.A. Meyers (Ed. 2008).

[13] A. A. Deniz, S. Mukhopadhyay, E. A. Lemke, *Single-molecule biophysics: at the interface of biology, physics and chemistry*, J. R. Soc. Interface, **5**, 15 - 45, (2008).

[14] T. Hoffman and L. Dougan ,*Single molecule force spectroscopy using polyproteins*, Chem. Soc. Rev, **41**, 4781 - 4796, (2012).

[15] A. Engel, H.E. Gaub, D.J. Muller, *Atomic force microscopy: a forceful way with single molecules*, Curr. Biol., **9**, R133-R136 (1999).

[16] M. Ludwig, M. Rief, L. Schmidt, H. Li, F. Oesterhelt, M. Gautel, H.E. Gaub, *AFM, a tool for single-molecule experiments*, Appl. Phys., **A68**, 173-176 (1999).

[17] T.E. Fisher, P.E. Marszalek, A.F. Oberhauser, M. Carrion-Vazquez, J.M. Fernandez, *The micro-mechanics of single molecules studied with atomic force microscopy*, J. Physiol. -London, **520**, 5-14, (1999)

[18] T.R. Strick, J.-F. Allemand, D. Bensimon, V. Croquette, *Stress-induced structural transitions in DNA and proteins*, Annu. Rev. Biophys. Biomed. Struct., **29**, 523-543, (2000).

[19] A. Janshoff, M. Neitzert, Y. Oberdörfer, H. Fuchs, *Force spectroscopy of molecular systems- single molecule spectroscopy of polymers and biomolecules*, Angew. Chem. Int. Ed., **39**, 3212-3237, (2000).

[20] G. Binnig, C.F. Quate, and C. Gerber, *Atomic force microscope*, Phys.Rev.Lett, **56**(6), 930-936, (1986).

[21] U. Maver, T. Maver, Z. Persin, K. Stana-Kleinschek, M. Mozetic and A. Vesel, *Polymer characterization with the atomic force microscope*, Polymer science, Chapter 4, 113 - 132, (2013).

[22] H. S. Wong, C. Durkan and N. Chandrasekar, *Tailoring the Local Interaction between Graphene Layers in Graphite at the Atomic Scale and Above Using Scanning Tunneling Microscopy*, ACS Nano,**3**, 11, 3455-3462, (2009).

[23] T. Ando, *High-speed atomic force microscopy*, Current opinion in Chemical Biology,**51**, 105-112, (2019).

[24] M. Owa, T. Uchihashu, H.A. Yanagisawa, T. Yamano, H. Iguchi, H. Fukuzawa, K.I. Wakabayashi, T. Ando and M. Kikkawa, *Inner lumen proteins stabilize doublet microtubules in cilia and flagella*, Nat. Communications,**1**, 1143, (2019).

[25] P. Kunda, A. E. Pelling, T. Liu and B Baum *Moesin controls cortical rigidity, cell rounding and spindle morphogenesisi during mitosis*, Current Biology,**18**, 91-101, (2008).

[26] M. Lekka, P. Laidler, D. Gil, J. Lekki, Z. Stachura and A. Z. Hrynkiewicz *Elasticity of normal and cancerous human bladder cells studied by scanning force microscopy*, European Biophysics Journal,**28**, 4, 312-316, (1999).

[27] I. Dulinska, M. Targosza, W. Strojnyc, M. Lekka, P. Czuba, W. Balwierz and M. Szymonski, *Stiffness of normal and pathological erythrocytes studied by means of atomic force microscopy*,J. Biochem. Biophys. Method,**66**, 1-11, (2006).

[28] M. Lekka ,D. Gil, K. Pogoda, J. Dulinska-Litewka, R. Jach, J. Gostek, O. Klymenko, S. Prauzner-Bechcicki, Z. Stachura,J. Wiltowska-Zuber, K. Okon and P. Laidler *Cancer cell detection in tissue sections using AFM*, Archives of Biochemistry and Biophysics,**518**, 151-156, (2012).

[29] M. Lekka, K. Pogoda, J. Gostek, O. Klymenko, S. Prauzner-Bechcicki, J. Jaczewska, J. Lekki and Z. Stachura, *Cancer cell recognition - Mechanical phenotype*, Micron,**43**, 12, 1259-1266, (2012).

[30] D. C. Lin, E. K. Dimitridas and F. Horkay *Robust strategies for automated AFM force curve analysis. Non-adhesive indentation of soft, inhomogeneous materials.*, ASME,**129**, 430-440, (2007).

[31] Y. Ding,G-K. Xu and G-F. Wang *On the determination of elastic moduli of cells by AFM based indentation*, Scientific Reports,**7**, 45575, (2017).

[32] JPK Instruments AG - Scanning Probe Technologies. *Determining the Elastic Properties of Biological Samples with AFM*, AZoNano, (2019).

[33] JPK Instruments Technical Note, *A practical guide to AFM force spectroscopy and data analysis*.

[34] R. W. Stark, T. Drobek and W. M. Heckl, *Thermomechanical noise of a free v-shaped cantilever for atomic-force microscopy*, Ultramicroscopy, **86**, 207-215 (2001).

[35] H-J. Butt and M. Jaschke, *Calculation of thermal noise in atomic force mycroscopy*, Nanotechnology,**6**, 1-7 (1995).

[36] F. Amblard, B. Yurke, A. Pargellis, and S. Leibler, *A magnetic manipulator for studying local rheology and micromechanical properties of biological systems*, Rev. Sci. Instrum., **67**, 818, (1996).

[37] T. R. Strick, J.-F. Allemand, D. Bensimon, and V. Croquette, *The Elasticity of a Single Supercoiled DNA Molecule*, Science, **271**, 1835-1837, (1996).

[38] H. Chen, G. Yuan, R. S. Winardhi, M. Yao, I. Popa, J. M. Fernandez and J. Yan, *Dynamics of Equilibrium Folding and Unfolding Transitions of TitinImmunoglobulin Domain under Constant Forces*, J. Am. Chem. Soc, **137**, 10, 3540-3546, (2015).

[39] I. D. Vilfan, J. Lipfert., D. A. Koster, S. G. Lemay, and N. H. Dekker, *Magnetic Tweezers for Single-Molecule Experiments*, in P. Hinterdorfer, A. Oijen (eds) Handbook of Single-Molecule Biophysics, Springer, 371-395, (2009).

[40] D. Salerno, D. Brogioli, V. Cassina, D. Turchi, G. L. Beretta, D. Seruggia, R. Ziano, F. Zunino, and F. Mantegazza, *Magnetic tweezers measurements of the nanomechanical properties of DNA in the presence of drugs*, Nucleic Acids Res., **38**, 7089-7099, (2010).

[41] D. Salerno, A. Tempestini, I. Mai, D. Brogioli, R. Ziano, V. Cassina, and F. Mantegazza, *Single-Molecule Study of the DNA Denaturation Phase Transition in the Force-Torsion Space*, Phys. Rev. Lett., **109**, 118303, (2012).

[42] A. Tempestini, V. Cassina, D. Brogioli, R. Ziano, S. Erba, R. Giovannoni, M. G. Cerrito, D. Salerno, and F. Mantegazza, *Magnetic tweezers measurements of the nanomechanical stability of DNA against denaturation at various conditions of pH and ionic strength*, Nucleic Acids Res., **43**, 2009-2012, (2013).

[43] J. Madariaga-Marcos, C. L Pastrana, G. L.M. Fisher, M. S. Dillingham and F. Moreno-Herrero, *ParB dynamics and the critical role of the CTD in DNA condensation unveiled by combined force-fluorescence measurements*, eLife, **8**, e43812, (2019).

[44] J. Madariaga-Marcos, S. Hormeno, C. L. Pastrana, G. L. M. Fisher, M. S. Dillingham and F. Moreno-Herrero, *Force determination in lateral magnetic tweezers combined with TIRF microscopy*, Nanoscale, **9**, 4579-4590, (2018).

[45] G. L. M. Fisher, C. L. Pastrana, V. A. Higman, A. Koh, J. A. Taylor, A. Butterer, T. Craggs, F. Sobott, H. Murray, M. P. Crump, F. Moreno-Herrero and M. S. Dillingham , *The structural basis for dynamic DNA binding and bridging interactions which condense the bacterial centromere*, eLife, **6**, e28086, (2017).

[46] C. L. Pastrana, C. Carrasco, P. Akhtar, S. h: Leuba, S. A. Khan and F. Moreno-Herrero, *Force and twist dependence of RepC nicking activity on torsionally-constrained DNA molecules*, Nucleic Acids Research, **44**, 18,8885-8896, (2016).

[47] J. Lipfert, X. Hao, and N. H. Dekker, *Quantitative Modeling and Optimization of Magnetic Tweezers*, Biophys. J., **96**, 5040-5049, (2010).

[48] C. Gosse and V. Croquette, *Magnetic Tweezers: Micromanipulation and Force Measurement at the Molecular Level*, Biophys. J., **82**, 3314-3329, (2002).

[49] I. D. Vilfan, J. Lipfert., D. A. Koster, S. G. Lemay, and N. H. Dekker, *Magnetic Tweezers for Single-Molecule Experiments*, in P. Hinterdorfer, A. Oijen (eds) Handbook of Single-Molecule Biophysics, Springer, 371-395, (2009).

[50] J. F. Marko and E. D. Siggia, *Stretching DNA*,Macromolecules, **26**, 8759–8770, (1998).

[51] J. H. Kim, W. X. Shi and R. G: Larson, *Methods of stretching DNA molecules using flow fields.*, Langmuir, **23**, 2, 775-62, (2007).

[52] S. Thakur, D. I Cattoni and M. Nollman *The fluorescence properties and binding mechanism of SYTOX green, a bright, low photo-damage DNA intercalating agent*, Eur. Biophys. J. , **44**, 337–348, (2015).

[53] M. Ganji, S. H. Kim, J. van der Torre, E. Abbondanzieri and C. Dekker, *Intercalation-Based Single-Molecule Fluorescence Assay To Study DNA Supercoil Dynamics.*, Nano Lett., **13**, 16(7), 4699-707, (2016).

[54] M. Ganji, I.A. Shaltiel, S. Bisht, E. Kim, A. Kalichava, C.H. Haering and C. Dekker, *Real-time imaging of DNA loop extrusion by condensin*, Science, **360**, 102-105, (2018).

[55] J. M. Eeftens and C. Dekker, *Catching DNA with hoops – biophysical approaches to clarify the mechanism of SMC proteins*, Nature Struct. Molec. Biol., **24**, 1012, (2017).

[56] J. van Mameren, E. J. Peterman and G. J. Wuite, *See me, feel me: methods to concurrently visualize and manipulate single DNA molecules and associated proteins*, Nucleic Acids Res, **36**, 4381 - 4389, (2008).

[57] S.H. Kim, M. Ganji, J. van der Torre, E. Abbondanzieri and C. Dekker, *DNA sequence encodes the position of DNA supercoils*, eLife, **7**, e36557, (2018).

[58] M. Ganji, S. H. Kim, J. van der Torre, E. Abbondanzieri and C. Dekker, *The condensin complex is a mechanochemical motor that translocates along DNA*, Science, eaan6516, (2017).

[59] C. S. Niman, J. P. Beech, J. O. Tegenfeldt, P. M- G. Curmi, D. N. Woolfson, N. R. Force and H. Linke, *Controlled microfluidic switching in arbitrary timesequences with low drag*, Lab on a Chip, **13**, 2389, (2013).

[60] G. Taylor, *Dispersion of soluble matter in solvent flowing slowly through a tube*, Proc. R. Soc. London, Ser. A, **219**,(1137), 186–203, (1953).

[61] R. Aris, *On the dispersion of a solute in a fluid flowing through a tube*, Proc. R. Soc. London, Ser. A, **235**,(1200), 67–77, (1956).

[62] J. Madariaga-Marcos, *Magnetic Tweezers and Fluorescence to study DNA:protein interactions*, PhD Thesis, Universidad Autonoma de Madrid, (2019).

[63] T. Wazawa and M. Ueda, *Total internal reflection fluorescence microscopy in single molecule nanobioscience.*,Adv Biochem Eng Biotechnol., **95**, 77-106,(2005).

[64] D. L. Nelson and M. Cox, *Lehninger principles of biochemistry*,W. H. Freeman; 4th edition, (2004).

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

[66] V. N. Uversky, *Intrinsically Disordered Proteins and Their "Mysterious" (Meta)Physics*, Front. Phys., **7**, 10, 1-18, (2019).

[67] 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*, Proc. Natl. Acad. Sci. USA, Biophysics, **96**, 3694-3699, (1999).

[68] Y. Cao, H. Li, *How do chemical denaturants affect the mechanical folding and unfolding of proteins?*, Journal of Molecular Biology,**375**(1), 316–324, (2008).

[69] J.E. Kohn, I. S. Millett, J. Jacob, B. Zagrovic, T.M. Dillon, N. Cingel, R.S. Dothager, S. Seifert, P. Thiyagarajan, T.R. Sosnick, M.Z. Hasan, V.S. Pande, I. Ruczinski, S. Doniach and K. W. Plaxco, *Random coil behavior and the dimensions of chemically unfolded proteins*, Proc. Natl. Acad. Sci. USA, Biophysics, **101**(34), 12491-12496, (2004).

[70] I.S. Millett, S. Doniach, K. W. Plaxco, *Towars a taxonomy of the denatured state: Small angle scattering studies of unfolded proteins*, Adv. Protein. Chem. **62**, 241-262, (2002).

[71] G. Stirnemann, K. Seung-gu, R. Zhou and B. J. Berne, *How force unfolding differs from chemical denaturation*, Proceedings of the National Academy of Sciences, **111** (9), 3413-3418, (2014).

[72] E. J. Guinn, B. Jagannathan and S. Marqusee, *Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein*, Nature Communications **6**, 6861, (2015).

[73] M. Carrion-Vazquez, A. F. Oberhauser, T. E. Fisher, P. E. Marszalek, H. Li and J. M. Fernandez, *Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineeringn*, Prog. Biophysics & Molecular Biology, **74**, 63-91, (2000).

[74] I. Popa, P. Kosuri, J. Alegre-Cebollada, S. Garcia-Manyes, J.M. Fernandez, *Force dependency of biochemical reactions measured by single-molecule force-clamp spectroscopy*, Nature Protocols, **8**, 1261-1276, (2013).

[75] H. Li, A.F. Oberhauser, S.B. Fowler, J. Clarke, J.M. Fernandez, *Atomic force microscopy reveals the mechanical design of a modular protein*, Proc. Natl. Acad. Sci. USA, **97** 6527–6531, (2000).

[76] M. Carrion-Vazquez, P.E. Marszalek, A.F. Oberhauser, J.M. Fernandez, *Atomic force microscopy captures length phenotypes in single proteins*, Proc. Natl. Acad. Sci. USA., **96**, 11288–11292, (1999).

[77] M. Rief, M. Gautel, F. Oesterhelt, J.M. Fernandez, H.E. Gaub, *Reversible unfolding of individual titin immunoglobulin domains by AFM*, Science, **276**, 1109–1112, (1997).

[78] A. S. Politou, D.J. Thomas and A. Pastore, *The folding and stability of Titin immunoglobulin-like modules, with implications for the mechanism of elasticity*, Biophysical Journal, **69**, 2601-2610, (1995).

[79] J. Alegre-Cebollada, P. Kosuri, D. Giganti, E. Eckels, J. A. Rivas-Pardo, N. Hamdani, C. M. Warren, R. J. Solaro, W. A. Linke

and J.M. Fernandez, *S-Glutathionylation of cryptic cysteines enhances Titin elasticity by blocking protein folding*, Cell, **156**, 1235-1246, (2014).

[80] S. Improta, A.S. Politou, A. Pastore, *Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity*, Structure, **4**, 323-337,(1996).

[81] L. Dougan and J.M. Fernandez, *Tandem repeating modular proteins avoid aggregation in single molecule force spectroscopy experiments*, The Journal of Physical Chemistry, **111**(49), 12402-8, (2008).

[82] A. K. Dunker, D. Lawson, C. J. Brown, R.M. Williams,.. E. C. Garner and Z. Obradovic, *Intrinsically disordered protein*, Journal of Moleculare Graphics and Modelling, **19**, 26-59, (2001).

[83] A. Klubera, T. A. Burta, and C. Clementia, *Size and topology modulate the effects of frustration in protein folding*, PNAS, **115**. 37, 9234–9239, (2018).

[84] C. Boesch, A. Bundi, M. Oppliger, and K. Wthrich, *1H nuclear-magneticresonance studies of the molecular conformation of monomeric glucagon in aqueous solution*, Eur. J. Biochem., **91**, 209–214,(1978).

[85] A.J. Daniels, R.J. Williams, and P.E. Wright. *The character of the stored molecules in chromaffin granules of the adrenal medulla: a nuclear magnetic resonance study.*, Neuroscience, **3**, 573–585, (1978).

[86] A. Schlessinger, C. Schaefer, E. Vicedo, M. Schmidberger, M. Punta and B. Rost, *Protein disorder — a breakthrough invention of evolution?*, Current Opinion in Structural Biology , **21**, 412–418, (2011).

[87] J. B. Ahrens, J. Nunez-Castilla and J. Siltberg-Liberles, *Evolution of intrinsic disorder in eukaryotic proteins*, Cell. Mol. Life Sci., **74**, 3163–3174, (2017).

[88] M. Sandal, F. Valle, I. Tessari, S. Mammi, E. Bergantino, F. Musiani, M. Brucale, L. Bubacco, B. Samorì, *Conformational Equilibria in monomeric α-synuclein at the single-molecule level*, Plos Biology, **6**, 99-108, (2008).

[89] M. Brucale, M. Sandal, s. Di Maio, A. Rampioni, I. Tessari, L. Tosatto, M. Bisaglia, L. Bubacco, B. Samorì, *Pathogenic mutations shift the equilibria of α-synuclein single molecules towards structured conformers*, ChemBioChem, **10**, 176-183, (2009).

[90] S.E. Salghetti, A.A. Caudy, J.G. Chenoweth, and W.P. Tansey. *Regulation of transcriptional activation domain function by ubiquitin.* Science, **293**, 1651– 1653, (2001).

[91] A.H. Huber and W.I. Weis. T*he structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by betacatenin.* Cell, **105**, 391–402, (2001).

[92] P. Borbat, T. F. Ramlall, J. H. Freed, D. Eliezer, I*nter-Helix Distances in Lysophospholipid Micelle-Bound α-Synuclein from Pulsed ESR Measurements* J. Am. Chem. Soc., **128**, (31),10004-10005, (2006).

[93] C. Wang ,C. Zhao, D. Li, Z. Tian, Y. Lai, J. Diao, and C. Liu, **Versatile Structures of α-Synuclein**, Front Mol Neurosci., **9**, 48, (2016).

[94] T. S. Ulmer, A. Bax, N. B. Cole, R. L. Nussbaum, *Structure and dynamics of micelle-bound human alpha-synuclein.*, J Biol Chem., **280**, 10, 595-603, (2005).

[95] V.N. Uversky. *Protein folding revisited. A polypeptide chain at the folding misfolding-nonfolding cross-roads: which way to go?* Cell. Mol. Life Sci., **60**, 1852–1871, (2003).

[96] A. Borgia, M.B. Borgia, K. Brugge, V. M. Kissling, P. O. Heidarsson, C. B. Fernandes, A. Sottini, A. Soranno, K. J. Buholzer, D. Nettels, B. B. Kragelund, R.B. Best, and B. Schuler,*Extreme disorder in an ultrahigh -affinity protein complex*, Nature, 5557**694**, 61-66, (2018).

[97] Y. Qu and D.W. Bolen. *Efficacy of macromolecular crowding in forcing proteins to fold.*, Biophysical Chemistry, **101**, 155–165, (2002).

[98] P. Tompa. *Intrinsically unstructured proteins.*, Trends in Biochemical Sciences, **27**(10),527–533, (2002).

[99] I. König, A. Zarrine-Afsar, M. Aznauryan, A. Soranno, B. Wunderlich, F. Dingfelder, J. C. Stüber, A. Plückthun, D. Nettels and B. Schuler, *Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells.* Nat Methods, 12, 773–779, (2015).

[100] F.-X. Theillet, A. Binolfi, B. Bekei, A. Martorana, H. M. Rose, M. Stuiver, S. Verzini, D. Lorenz, M. van Rossum, D. Goldfarb and P. Selenko, *Structural disorder of monomeric α-synuclein persists in mammalian cells* Nature, 530, 45-50, (2016).

[101] M. Goederta, R. Jakesa, M. G. Spillantini, *The Synucleinopathies: Twenty Years On*, Journal of Parkinson's Disease, **7**, S51-S69, (2017).

[102] L. Maroteaux, J. T. Campanelli, R. H. Scheller, *Synuclein: a neuron-specific protein localized to the nucleus and presynaptic nerve terminal*, Journal of Neuroscience, **8**, 2804-2815, (1988).

[103] P. H. Weinreb, W. Zhen, A. W. Poon, K. A. Conway, P. T. Lansbury Jr., *NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded*, Biochemistry, **35**, 13709-13715, (1996).

[104] I. V. J. Murray, Virginia M.-Y. Lee, J. Q. Trojanowski, *Synucleinopathies: a pathological and molecular review*, Clinical Neuroscience Research, **1**, 445-455, (2001).

[105] M. Bisaglia, S. Mammi, L. Bubacco, *Structural insights on physiological functions and pathological effects of α-synuclein*, The FASEB Journal, **23**, 330, (2009).

[106] T. Bartels, L. S. Ahlstrom, A. Leftin, F. Kamp, C. Haass, M. F. Brown, K. Beyer, *The N-terminus of the intrinsically disordered protein α-synuclein triggers membrane binding and helix folding*, Biophysical Journal, **99**, 2116-2124, (2010).

[107] R. Krüger, W. Kuhn, T. Müller, D. Woitalla, M. Graeber, S. K osel, H. Przuntek, J. T. Epplen, L. Sch ols, O. Riess, *Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease*, Nature Genetics, **18**, 106-108, (1998).

[108] J. J. Zarranz, J. Alegre, J. C. Gómez-Esteban, E. Lezcano, R. Ros, I. Ampuero, L. Vidal, J. Hoenicka, O. Rodriguez, B. Atarés, V. Llorens, E. Gomez Tortosa, T. del Ser, D. G. Muñoz, J. G. de Yebenes, *The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia.*, Annals of Neurology, **55**, 164-173, (2004).

[109] M. H. Polymeropoulos, C. Lavedan,E. Leroy, .., L. I. Golbe, R. L. Nussbaum, *Mutation in the alpha-synuclein gene identified in families with Parkinson's disease*, Science, **276**, 2045-2047, (1997).

[110] S. Appel-Cresswell, C. Vilarino-Guell, M. Encarnacion, H. Sherman, I. Yu, B. Shah, D. Weir, C. Thompson, C. Szu-Tu, J. Trinh, J. O. Aasly, A. Rajput, A. H. Rajput, A. J. Stoessl, M. J. Farrer, *Alpha-Synuclein p.H50Q, a Novel Pathogenic Mutation for Parkinson's Disease*, Movement Disorders, **28**, (2013).

[111] S. Lesage, M. Anheim, F. Letournel, L. Bousset, A. Honore, N. Rozas, L. Pieri, K. Madiona, A. Durr, R. Melki, C. Verny, A. Brice, *G51D alpha-synuclein mutation causes a novel Parkinsonian–pyramidal syndrome*, Annals of Neurology, **73**, 459-471, (2013).

[112] B. I. Giasson, I. V. Murray, J. Q. Trojanowski, V. M. Lee, *A hydrophobic stretch of 12 amino acid residues in the middle of alpha-synuclein is essential for filament assembly*, Jornal of Biological Chemistry, **276**, 2380-2386, (2001).

[113] E. A. Waxman, K. L. Emmer, B. I. Giasson, *Residue Glu83 plays a major role in negatively regulating α-synuclein amyloid formation*, Biochemical and biophysical research communications, **391**, 1415-1420, (2010)

[114] I. V. Murray, B. I. Giasson, S. M. Quinn, V. Koppaka, P. H. Axelsen, H. Ischiropoulos, J. Q. Trojanowski, V. M. Lee, *Role of alpha-synuclein carboxy-terminus on fibril formation in vitro*, Biochemistry, **42**, 8530-8540, (2003).

[115] C. Lavedan, *The synuclein family*, Genome Res, **8**, 871-880, (1998).

[116] K. Beyer, *Alpha-synuclein structure, posttranslational modification and alternative splicing as aggregation enhancers*, Acta Neuropathologica, **112**, 237-251, (2006).

[117] B. Fauvet, M. K. Mbefo, M.B. Fares, C. Desobry, ..., E. Masliah, and H.A. Lashuel, *α-synuclein in central nervous system and from erythrocytes, mammalian cells, and Escherichia coli exists predominantly as disordered monomer.*, J Biol Chem., **287**, (19), 15345-64, (2012).

[118] R. Bussell, and D. Eliezer, *Residual Structure and Dynamics in Parkinson's Disease-associated Mutants of α-Synuclein*, The Journal of Biological Chemistry, **276**, (5), 45996-46003, (2001).

[119] U. Dettmer, A. J. Newman, E. S. Luth, T. Bartels and D. Selkoer, *In Vivo Cross-linking Reveals Principally Oligomeric Forms of α-Synuclein and β-Synuclein in Neurons and Non-neural Cells*, The Journal of Biological Chemistry, **288**, (5), 6371-6385, (2013).

[120] J. Burre, M. Sharma, and T. C. Sudhof, *α-Synuclein assembles into higher-order multimers upon membrane binding to promote SNARE complex formation*, PNAS, **11**, (40), E4274-E4283, (2014).

[121] K. K. Dev, K. Hofele, S. Barbieri, V. I. Buchman, H. van der Putten, *Alpha-synuclein and its molecular pathophysiological role in neurodegenerative disease*, Neuropharmacology, **45**, 14-44, (2003).

[122] A. Natalello, F. Benetti, S. M. Doglia, G. Legname, and R. Grandori, *Compact conformations of a-synuclein induced by alcohols and copper*, Proteins, **79**, 611–621, (2010).

[123] J. C. Watts, A. Tandon and P. E. Fraser, *The biology an pathobiology of α-synuclein*, Protein Folding Disorders of the central nervous system, Proteins, **5**, 109-130, (2017).

[124] A. Iway, E. Masliah, M. Yoshimoto, N. Ge, L. Flanagan, H. A. R. De Silva, A. Kittel, T. Saitoh, *The precursor protein of non-Aβ component of Alzheimer's disease amyloid is a presynaptic protein of the central nervous system*, Neurons, **14**, 467-475, (1995).

[125] V. N. Uversky, and D. Eliezer, *Biophysics of Parkinson's disease: structure and aggregation of alpha-synuclein.*, Curr Protein Pept Sci, **10**, (5), 483-995, (2009).

[126] M. Necula, C. N. Chirita, J. Kuret, *Rapid anionic micelle-mediated alpha-synuclein fibrillization in vitro*, Journal of Biological Chemistry, **278**, 46674-44680, (2003).

[127] M. M. Dedmon, K. Lindorff-Larsen, J. Christodoulou, M. Vendruscolo, C. M. Dobson,*Mapping long-range interactions in alphasynuclein using spin-label NMR and ensemble molecular dynamics simulations*, Journal of the American Chemical Society, **127**, 476-477, (2005).

[128] C. W. Bertoncini, Y. S. Jung, C. O. Fernandez, W. Hoyer, C. Griesinger, T. M. Jovin, M. Zweckstetter, *Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein*, Proceedings of the National Academy of Sciences, **102**, 1430-1435, (2005).

[129] J. R. Allison, P. Varnai, C. M. Dobson, M. Vendruscolo, *Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements*, Journal of the American Chemical Society, **131**, 18314-18326, (2009).

[130] T. L. Yap, C. M. Pfefferkorn, J. C. Lee, *Residue-specific fluorescent probes of α-synuclein: detection of early events at the N- and C-termini during fibril assembly*, Biochemistry, **50**, 1963-1965, (2011).

[131] A. J. Trexler, E. Rhoades, *Single molecule characterization of α-synuclein in aggregation-prone states*, Biophysical Journal, **99**, 3048-3055, (2010).

[132] T. Takahashi, H. Mihara, *Peptide and protein mimetics inhibiting amyloid beta-peptide aggregation*, Accounts of Chemical Research, **41**, 1309-1318, (2008).

[133] B. Winner, R. Jappelli, S. K. Maji, P. A. Desplats, L. Boyer, S. Aigner, C. Hetzer, T. Loher, M. Vilar, S. Campioni, C. Tzitzilonis, A. Soragni, S. Jessberger, H. Mira, A. Consiglio, E. Pham, E. Masliah, F. H. Gage, R. Riek, *In vivo demonstration that alpha-synuclein oligomers are toxic*, Proceedings of the National Academy of Sciences, **108**, 4194-4199, (2011).

[134] K. A. Conway, S. J. Lee, J. C. Rochet, T. T. Ding, R. E. Williamson, P. T. Lansbury, *Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset Parkinson's disease: implications for pathogenesis and therapy*, Proceedings of the National Academy of Sciences, **97**, 571-576, (2000).

[135] R. A. Fredenburg, C. Rospigliosi, R. K. Meray, J. C. Kessler, H. A. Lashuel, D. Eliezer, P. T. Lansbury, *The impact of the E46K mutation on the properties of alpha-synuclein in its monomeric and oligomeric states*, Biochemistry, **46**, 7107-7118, (2007).

[136] L. Giehm, D. I. Svergun, D. E. Otzen, B. Vestergaard, *Low-resolution structure of a vesicle disrupting alpha-synuclein oligomer that accumulates during fibrillation*, Proceedings of the National Academy of Sciences, **108**, 3246-3251, (2011).

[137] Y. C. Wong, D. Krainc, *α-synuclein toxicity in neurodegeneration: mechanism and therapeutic strategies*, Nature Medicine, **23**, (2017).

[138] S. J. Wood, J. Wypych, S. Steavenson, J. C. Louis, M. Citron, A. L. Biere, *Alpha-synuclein fibrillogenesis is nucleation-dependent. Implications for the pathogenesis of Parkinson's disease*, Journal of Biological Chemistry, **274**, 19509-12, (1999).

[139] F. N. Emamzadeh, *Alpha-synuclein structure, functions, and interactions*, Journal of Research in Medical Sciences, **21**, (2016).

[140] Leonidas Stefanis, *α-Synuclein in Parkinson's Disease*, Cold Spring Harbor Perspectives in Medicine, **4**, (2012).

[141] J. Wang, X. Han, N. A. Leu, S. Sterling, S. Kurosaka, M. Fina, V. M. Lee, D. W. Dong, J. R. Yates, A. Kashina, *Protein arginylation targets alpha synuclein, facilitates normal brain health, and prevents neurodegeneration*, Scientific Reports, **7**, (2017).

[142] J. Renaud, S. F. Nabavi, M. Daglia, S. M. Nabavi, M. G. Martinoli1, *Epigallocatechin-3-Gallate, a Promising Molecule for Parkinson's Disease?*, Rejuvenation Research, **18**, (2015).

[143] J. Zhao, Q. Liang, Q. Sun, C. Chen, L. Xu, Y. Ding, P. Zhou, *Epigallocatechin-3-gallate (EGCG) inhibits fibrillation, disaggregates amyloid fibrils of α-synuclein, and protects PC12 cells against α-synuclein-induced toxicity*, RSC Advances, **7**, 32508-32517, (2017).

[144] X. Liu, S. Zhou, D. Shi, Q. Bai, H. Liu, X. Yao, *Influence of EGCG on α-synuclein (αS) aggregation and identification of their possible binding mode: A computational study using molecular dynamics simulation*, Chemical Biology and Drug Design, **91**, 162-171, (2018).

[145] F. T. Outeiro, J. Klucken1, K. Bercury, J. Tetzlaff, P. Putcha, L. M. A. Oliveira, A. Quintas, P. J. McLean, B. T. Hyman, *Dopamine-Induced Conformational Changes in Alpha-Synuclein*, Plos One, **4**, (2009).

[146] H. J. Lee, S. M. Baek, D. H. Ho, J. E. Suk, E. D. Cho, S. J. Lee, *Dopamine promotes formation and secretion of non-fibrillar alpha-synuclein oligomers*, Experimental and Molecular Medicine, **43**, 216-222, (2011).

[147] E. Illes-Toth, C. F. Dalton, D. p. Smith, *Binding of Dopamine to alpha-synuclein is mediated by specific conformational states*, Journal of the American Society Mass Spectrometry, **24**, 1346-1354, (2013).

[148] F. E. Herrera, A. Chesi, K. E. Paleologou, A. Schmid, A. Munoz, M. Vendruscolo, S. Gustincich, H. A. Lashuel, P. Carloni, *Inhibition of α-Synuclein fibrillization by dopamine is mediated by interactions with five C-terminal residues and with E83 in the NAC region*, Plos One, **3**, (2008).

[149] A. Konijnenberg, S. Ranica, J. Narkiewicz, G. Legname, R. Grandori, F. Sobott, A. Natalello, *Opposite Structural Effects of Epigallocatechin-3-gallate and Dopamine Binding to α-Synuclein.*, Anal. Chem., **88**, 8468-8475, (2016).

[150] T. R. Alderson and J. L. Markley, *Biophysical characterization of α-synuclein and its controversial structure*, Intrinsically disordered proteins, **1**, (1), e26255, (2013).

[151] P. H. Weinreb, W. Zhen, A. W. Poon, K. A. Conway, and P. T. Lansbury, *NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded.*, Biochemistry, **35**, 13709-15, (1996).

[152] K. Gast, H. Damaschun, K. Eckert, K. Schulze-Forster, H. R. Maurer, M. Muller-Frohne, D. Zirwer, J. Czarnecki, G. Damaschun, *Prothymosin alpha: a biologically active protein with random coil conformation.*, Biochemistry, **34**, 13211-8, (1995).

[153] V. N. Uversky, J. Li, A. L. Fink, *Evidence for a partially folded intermediate in alpha-synuclein fibril formation.*, J Biol Chem, **276**, 10737-44, (2001).

[154] C. W. Bertoncini, Y. S. Jung, C. O. Fernandez, W. Hoyer, C. Griesinger, T. M. Jovin, M. Zweckstetter, R*elease of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein.* Proc Natl Acad Sci, **102**, 1430-5, (2005).

[155] P. Bernadó, C. W. Bertoncini, C. Griesinger, M. Zweckstetter, M. Blackledge, *Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings.*, J Am Chem Soc, **127**, 17968-9, (2005).

[156] A. D. Stephens, M. Zacharopoulou, G. S. Kaminski Schierle, *The Cellular Environment Affects Monomeric α-Synuclein Structure.*, Trends Biochem. Sci. , **44**, 453–466, (2019).

[157] C. Santambrogio, A. Natalello, S. Brocca, E. Ponzini, R. Grandori, *Conformational Characterization and Classification of Intrinsically Disordered Proteins by Native Mass Spectrometry and*

Charge-State Distribution Analysis, Proteomics , **19**, (6), 1800060, (2019).

[158] P. Wongkongkathep, J.Y. Han, T.S. Choi, S. Yin, H.I. Kim, J. A. Loo, *Native Top-Down Mass Spectrometry and Ion Mobility MS for Characterizing the Cobalt and Manganese Metal Binding of α-Synuclein Protein.*, J. Am. Soc. Mass Spectrom., **29**, 1870–1880, (2018).

[159] L. Testa, S. Brocca, C. Santambrogio, A. D'Urzo, J. Habchi, S. Longhi, V. N. Uversky, R. Grandori, E*xtracting structural information from charge-state distributions of intrinsically disordered proteins by non-denaturing electrospray-ionization mass spectrometry.*, Intrinsic. Disord. Proteins, **1**, 25068, (2013).

[160] A. K. Frimpong, R. R. Abzalimov, V. N. Uversky, I. A. Kaltashov, *Characterization of intrinsically disordered proteins with electrospray ionization mass spectrometry: conformational heterogeneity of α-synuclein.*, Proteins, **78**, 714-722, (2010).

[161] U. H. Verkerk, P. Kebarle, *Ion-ion and ion-molecule reactions at the surface of proteins produced by nanospray. Information on the number of acidic residues and control of the number of ionized acidic and basic residues.*, J. Am. Soc. Mass Spectrom., **16**, 1325–1341, (2015).

[162] E. Ponzini, A. De Palma, L. Cerboni, A. Natalello, R. Rossi, R. Moons, A. Konijnenberg, J. Narkiewicz, G. Legname, F. Sobott, P. Mauri, C. Santambrogio and R. Grandori, *Methionine oxidation in α-synuclein inhibits its propensity for ordered secondary structure.*, J. Biol. Chem., **294**, 5657-5665, (2019).

[163] S. Nath, J. Meuvis, J. Hendrix, S. A. Carl, and Y. Engelborghs, *Early Aggregation Steps in α-Synuclein as Measured by FCS and FRET: Evidence for a Contagious Conformational Change*, Biophys J., **98**, (7) 1302–1311, (2010).

[164] M. Brucale, M. Schuler and B. Samorì *Single-Molecule Studies of Intrinsically Disordered Proteins*, Chem Rev., **114**,3281-3317, (2014).

[165] M. Schuler, H. Hoffman, *Single-molecule spectroscopy of protein folding dynamics–expanding scope and timescales.*, Curr. Opin. Struct. Biol., **23**, (1), 36-47, (2013).

[166] A. C. Ferreon, C. R. Moran, Y. Gambin, A. A. Deniz ,(2010) *Single-molecule fluorescence studies of intrinsically disordered proteins.* Methods Enzymol, **472**, 179–204, (2010).

[167] L. Tosatto, M. H. Horrocks, A. J. Dear, T. P. J. Knowles, M. Dalla Serra, N. Cremades, C. M. Dobson, and D. Klenermana, *Single-*

molecule FRET studies on alpha-synuclein oligomerization of Parkinson's disease genetically related mutants, Sci Rep., **5**, 16696, (2015).

[168] A. J. Trexler, E. Rhoades, *Single Molecule Characterization of α-Synuclein in Aggregation-Prone States*, Biophysical Journal, **99**, (9), 3048-3055, (2010).

[169] R. Hervás, J. Oroz, A. Galera-Prat, O. Goñi, A. Valbuena, A. M. Vera, A. Gómez-Sicilia, F. Losada-Urzáiz, V. N. Uversky, M. Menéndez, D.V. Laurents, M. Bruix, M. Carrión-Vázquez, *Common features at the start of the neurodegeneration cascade.*, PLOS Biol., **10**, 1001335, (2016).

[170] S. Xie, *Single-molecule approach to enzymology.* Single Mol., **4**, 229–236, (2001).

[171] M. Fernández-Ramírez, R. Hervás, A. Galera-Prat, V. Laurents, M. Carrión-Vázquez,*Efficient and simplified nanomechanical analysis of intrinsically disordered proteins*, Nanoscale, **10**, 216857, (2018).

[172] K. Pauwels, P. Lebrun, P. Tompa, *To be disordered or not to be disordered: is that still a question for proteins in the cell?*, Cell Mol Life Sci., **74**, (17),3185–3204,(2017).

[173] O. Coskuner-Weber, and V. N. Uversky, *Insights into the Molecular Mechanisms of Alzheimer's and Parkinson's Diseases with Molecular Simulations: Understanding the Roles of Artificial and Pathological Missense Mutations in Intrinsically Disordered Proteins Related to Pathology*, Int. J. Mol. Sci., **19**, 336, (2018).

[174] N. I. Brodie, K. I. Popov, E. V. Petrotchenko, N. V. Dokholyan, C. H. Brochers, *Conformational ensemble of native α-synuclein in solution as determined by short-distance crosslinking constraint-guided discrete molecular dynamics simulations*, PLOS Computational Biology, **15**,(3), e1006859, (2019).

[175] J.D. Watson and F.H.C. Crick, *A Structure for Deoxyribose Nucleic Acid*, Nature, **4356**, 737-738, (1953).

[176] T. R. Strick, M.-N. Dessinges, G. Charvin, N. H. Dekker, J.-F. Allemand, D. Bensimon, and V. Croquette, *Stretching of macromolecules and proteins*, Rep. Prog. Phys., **66**, 1-45, (2003).

[177] W. Kabsch, C. Sander and E. N. Trifonov, *The ten helical twist angles of B-DNA*, Nucleic Acids Res., **10**, 1097-1104, (1982).

[178] C. R. Calladine, H. R. Drew, B. F. Luisi, and A. A. Travers, *Understanding DNA*, Elsevier Academic Press, San Diego, (2004).

[179] A. Vologodskii, *Biophysics of DNA*, Cambridge University Press, Cambridge, (2015).

[180] J. C. Wang, *Helical repeat of DNA in solution*, Proc. Natl. Acad. Sci., **76**, 200-203, (1979).

[181] R. E. Dickerson, and H.-L. Ng, *DNA structure from A to B*, Proc. Natl. Acad. Sci., **98**, 6986-6988, (2001).

[182] P. L. Privalov, A. I. Dragan, C. Crane-Robinson, K. J. Breslauer, D. P. Remeta, and C. A. S. A. Minetti, *What Drives Proteins into the Major or Minor Grooves of DNA?*, J. Mol. Biol., **365**, 1-9, (2007).

[183] F. Uhllman, *SMC complexes: from DNA to chromosomes*, Nature Reviews Mol. Cell Bio., **17**, 399-412, (2016).

[184] H. Niki, A. Jeffè, R. Amamura, T. Ogura and S. Hiraga *The new gene mukB codes for a 177 kd protein with coiled-coiled domains involved in chromosome partitioning of E. coli*, EMBO J., **10**, 183-193, (1991).

[185] H. Niki, R. Imamura, M. Kitaoka, T. Ogura and S. Hiraga *E. coli MukB protein involved in chromosome partition forms a homodimer with a rodand-hinge structure having DNA binding and ATP/GTP binding activities.*, EMBO J., **11**, 5101–5109, (1992).

[186] C. Michaelis, R. Ciosk and K. Nasmyth, *Cohesins: chromosomal proteins that prevent premature separation of sister chromatids.*, Cell, **91**, 35-45, (1997).

[187] V. Guacci, D. Koshland and A. Strunnikov, *A direct link between sister chromatid cohesion and chromosome condensation revealed through analysis of MCD1 in S. cerevisiae*, Cell, **91**, 47-57, (1997).

[188] M. I. Fousteri and A. R. Lehmann, *A novel SMC protein complex in Schizosaccharomyces pombe contains the Rad18 DNA repair complex*, EMBO J., **19**,1691–1702 (2000).

[189] A. Tedeschi, .. and J. M. Peters, *Wapl is an essential regulator of chromatin structure and chromosome segregation.*, Nature, **501**,564-568 (2013).

[190] R. P. Birkenbihl and S. Subramani, *Cloning and characterization of rad21 an essential gene of Schizosaccharomyces pombe involved in DNA double-strand-break repair*, Nucleic Acids Res., **20**, 6605–6611 (1992).

[191] C. H. Haering, J. Lowe, A. Hochwagen,and K. Nasmyth, *Molecular architecture of SMC proteins and the yeast cohesin complex*, Mol. Cell **9**, 773–788 (2002).

[192] A. Lammens, A. Schele and K.P. Hopfner, *Structural biochemistry of ATP-driven dimerization and DNA stimulated activation of SMC ATPases*. Curr. Biol., **14**, 1778–1782 (2004).

[193] J. Palecek, S. Vidot, M. Feng, A. J. Doherty, and A. R. Lehmann, *The Smc5-Smc6 DNA repair complex. Bridging of the Smc5-Smc6 heads by the kleisin, Nse4, and non-kleisin subunits.*, J. Biol. Chem., **281**, 36952–36959 (2006).

[194] I. Onn, N. Aono, M. Hiranoand T Hirano, *Reconstitution and subunit geometry of human condensin complexes.*, EMBO J., **26**, 1024–1034,(2007).

[195] P. J. Huis et al., *Characterization of a DNA exit gate in the human cohesin ring.* Science, **346**, 968–972 (2014).

[196] L. Aragon, E. Martinez-Perez and M. Merkenschlager M, *Condensin, cohesin and the control of chromatin states.* Curr Opin Genet Dev, **23**, 204–211,(2013).

[197] D. E. Anderson, A. Losada, H. P. Erickson and T. Hirano, *Condensin and cohesin display different arm conformations with characteristic hinge angles.* J Cell Biol, **156**, 419–424, (2002).

[198] J. M. Eeftens, A. J. Katan, M. Kschonsak, M. Hassler, L. de Wilde, E. M. Dief, C. H. Haering and C. Dekker, *Condensin Smc2-Smc4 dimers are flexible and dynamic.*, Cell Rep **14**, 1813–1818, (2016).

[199] M. Hassler, I. A. Shaltiel and C. H. Haering, *Towards a unified model of SMC complec function*, Current Biology, **28**, R1266-R1281, (2018).

[200] T. M.K. Cheng, S. Heeger, R. A. G. Chaleil, N. Matthews, A. Stewart, J. Wright, C. Lim, P. A. Bates and F. Uhlmann, *A simple biophysical model emulates budding yeast chromosome condensation* eLife, **4**, e05565, (2015).

[201] F. Bürmann, B.G. Lee, T. Than, L. Sinn, F. J. O'Reilly, S. Yatskevich, J. Rappsilber, B. Hu, K. Nasmyth and J Löwe, *A folded conformation of MukBEF and cohesin*, Nature Structural and Molecular Biology, **26**, 227–236, (2019).

[202] P. Gutierrez-Escribano, M. D. Newton, A, Llauró, J. Huber, L. Tanasie, J. Davy, I. Aly, R. Aramayo, A. Montoya, H. Kramer, J. Stigler, D. S. Rueda and L. Aragon, *A conserved activity for cohesin in bridging DNA molecules*, Preprint, doi 10.1101/757286., (2019).

[203] N. Rezaei-Ghaleh, G. Parigi, A. Soranno, A. Holla, S. Becker, B. Schuler, C. Luchinat, and M. Zweckstetter, *Local and Global Dynamics in Intrinsically Disordered Synuclein*, Angew. Chem. Int. Ed. Engl., **57**, 15262-15266, (2018).

[204] A. Borgia, K.R. Kemplen, M.B. Borgia, A. Soranno, S. Shammas, B. Wunderlich, D. Nettels, R.B. Best, J. Clarke and B. Schuler, *Transient misfolding dominates multidomain protein folding.*, Nat. Commun., **6**, 8861, (2015).

[205] M. Gruebele, K. Dave, and S. Sukenik, *Globular Protein Folding In Vitro and In Vivo*. Annu. Rev. Biophys., **45**, 233-251, (2016).

[206] P.E. Wright, H.J. Dyson, *Intrinsically disordered proteins in cellular signalling and regulation.*, Nat. Rev. Mol. Cell. Biol., **16**, 18-29, (2015).

[207] A. F. Oberhauser, P. E. Marszalek, M. Carrion-Vazquez, and J. M. Fernandez, *Single protein misfolding events captured by atomic force microscopy.*, Nat. Struct. Bio., **6**, 102510-28, (1999).

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

[209] A.C.M. Ferreon, A. A. Deniz, *Protein folding at single-molecule resolution.* Biochim. Biophys. Acta, **1814**, 1021-1029, (2011).

[210] J.P. Junker, M. Rief, S*ingle-molecule force spectroscopy distinguishes target binding modes of calmodulin*. Proc. Natl. Acad. U.S.A., **106**, 14361-14366, (2009).

[211] A. D. Stephens, M. Zacharopoulou, G. S. Kaminski Schierle, *The Cellular Environment Affects Monomeric α-Synuclein Structure.*, Trends Biochem. Sci., **44**, 453-466, (2016).

[212] C. C. Curtain, N. M. Kirby, H. D. Mertens, K. J. Barnham, R. B. Knott, C. L. Masters, R. Cappai, A. Rekas, V. B. Kenche, and T. Ryan, *α-synuclein oligomers and fibrils originate in two distinct conformer pools: a small angle X-ray scattering and ensemble optimisation modelling study.* Mol. Biosyst., **11**, 190-196, (2015).

[213] J. Zhao, Q. Liang, Q. Sun, C. Chen, L. Xu, Y. Ding, P. Zhou, *Epigallocatechin-3-gallate (EGCG) inhibits fibrillation, disaggregates amyloid fibrils of α-synuclein, and protects PC12 cells against α-synuclein-induced toxicity*, RSC Advances, **7**, 32508-32517, (2017).

[214] H. J. Lee, S. M. Baek, D. H. Ho, J. E. Suk, E. D. Cho, S. J. Lee, *Dopamine promotes formation and secretion of non-fibrillar alpha-synuclein oligomers*, Experimental and Molecular Medicine, **43**, 216-222, (2011).

[215] A. Steward, J. L. Toca-Herrera, J. Clarke, *Versatile cloning system for construction of multimeric proteins for use in atomic force microscopy*, Protein science, **11**, 2179-2183, (2002).

[216] H. Jeong, V. Barbe, C. H. Lee, D. Vallenet, D. S. Yu, S.-H. Choi, A. Couloux, S.-W. Lee, S. H. Yoon, L. Cattolico, C.-G. Hur, H.-S. Park, B. Segurens, S. C. Kim, T. K. Oh, R. E. Lenski, F. W. Studier, P. Daegelen, J. F. Kim, *Genome Sequences of Escherichia coli B strains REL606 and BL21 (DE3)*, Journal of Molecular Biology, **394**, 644-652, (2009).

[217] A. Marbach, K. Bettenbrock, *lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA*, Journal of Biotechnology, **157**, 82-88, (2012).

[218] D. Latawiec, F. Herrera, A. Bek, V. Losasso, M. Candotti, F. Benetti, E. Carlino, A. Kranjc, M. Lazzarino, S. Gustincich, P. Carloni, and G. Legname, *Modulation of alpha-synuclein aggregation by dopamine analogs.*, PLoS One, **5**,(2), e9234, (2010).

[219] R. B. Best, D. J. Brockwell, J.L. Toca-Herrera, A. W. Blake, A. Smith, S. E. Radford, J. Clarke, *Force mode atomic force microscopy as a tool for protein folding studies*, J. Anal. Chim. Acta, **479**, 87-105, (2003).

[220] F. T. Outeiro, J. Klucken1, K. Bercury, J. Tetzlaff, P. Putcha, L. M. A. Oliveira, A. Quintas, P. J. McLean, B. T. Hyman, *Dopamine-Induced Conformational Changes in Alpha-Synuclein*, Plos One, **4**, (2009).

[221] K. A. Conway, S. J. Lee, J. C. Rochet, T. T. Ding, R. E. Williamson, P. T. Lansbury, *Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset Parkinson's disease: implications for pathogenesis and therapy*, Proceedings of the National Academy of Sciences, **97**, 571-576, (2000).

[222] L. Stefanis, *α-Synuclein in Parkinson's Disease*, Cold Spring Harbor Perspectives in Medicine, **4**, (2012).

[223] G. T. Heller, M. Bonomi, M. Vendruscolo, *Structural Ensemble Modulation upon Small-Molecule Binding to Disordered Proteins.*, J. Mol. Biol., **430**, 2288-2292, (2018).

[224] G. Rossetti, F. Musiani, E. Abad, D. Dibenedetto, H. Mouhib, C. O. Fernandez, P. Carloni, *Conformational ensemble of human α-synuclein physiological form predicted by molecular simulations*, Phys. Chem. Chem. Phys., **18**, 5702-5706, (2016).

[225] A. Balupuri, K. E. Choi, N. S. Kang, *Computational insights into the role of α-strand/sheet in aggregation of α-synuclein*, Sci Rep., **9**, 59, (2019).

[226] E. Smith, M. E. Jones, and P. A. Drew,*Quantitation of DNA methylation by melt curve analysis*, BMC Cancer, **9**, 123, (2009).

[227] L. N. Sanford, J. O. Kent, and C. T. Wittwer, *Quantum Method for Fluorescence Background Removal in DNA Melting Analysis*, Anal. Chem., **85**, 9907-9915, (2013).

[228] D. Salerno, D. Brogioli, V. Cassina, D. Turchi, G. L. Beretta, D. Seruggia, R. Ziano, F. Zunino, and F. Mantegazza, *Magnetic tweezers measurements of the nanomechanical properties of DNA in the presence of drugs*, Nucleic Acids Res., **38**, 7089-7099, (2010).

[229] A. Tempestini, V. Cassina, D. Brogioli, R. Ziano, S. Erba, R. Giovannoni, M. G. Cerrito, D. Salerno, and F. Mantegazza, *Magnetic tweezers measurements of the nanomechanical stability of DNA against denaturation at various conditions of pH and ionic strength*, Nucleic Acids Res., **43**, 2009-2012, (2013).

[230] V. Cassina, M. Manghi, D. Salerno, A. Tempestini, V. Iadarola, L. Nardo, S. Brioschi, and F. Mantegazza, *Effects of cytosine methylation on DNA morphology: An atomic force microscopy study*, Biochim. Biophys. Acta - Gen. Subj., **1860**, 1-7, (2016).

[231] V. Cassina, D. Seruggia, G. L. Beretta, D. Salerno, D. Brogioli, S. Manzini, F. Zunino, and F. Mantegazza, *Atomic force microscopy study of DNA conformation in the presence of drugs*, Eur. Biophys. J., **40**, 59-68, (2011).

[232] P. Sulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. Doye, and A. A. Louis, *Sequence-dependent thermodynamics of a coarse-grained DNA model*. J. Chem. Phys., **137**, 135101, (2012).

[233] A. Sen, and P. E. Nielsen, *Hydrogen bonding versus stacking stabilization by modified nucleobases incorporated in PNA·DNA duplexes*, Biophys. Chem., **141**, 29-33, (2009).

[234] J. SantaLucia, *A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics*, Proc. Natl. Acad. Sci., **95**, 1460-1465, (1998).

[235] J. SantaLucia, and D. Hicks, *The Thermodynamics of DNA Structural Motifs*, Annu. Rev. Biophys. Biomol. Struct., **33**, 415-440. (2004).

[236] J. Virstedt, T. Berge, R. M. Henderson, M. J. Waring, and A. A. Travers, *The influence of DNA stiffness upon nucleosome formation*, J. Struct. Biol., **148**, 66-85, (2004).

[237] A. Podestà, M. Indrieri, D. Brogioli, G. S. Manning, P. Milani, R. Guerra, L. Finzi, and D. Dunlap, *Positively charged surfaces increase the flexibility of DNA*, Biophys. J., **89**, 2558-2563, (2005).

[238] P. Wiggins, T. V. Der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson, *High flexibility of dna on short length scales probed by atomic force microscopy*, Nat. Nanotechnol., **1**, 137-141, (2009).

[239] C. Rivetti, M. Guthold, and C. Bustamante, *Scanning force microscopy of DNA deposited onto mica: Equilibration versus kinetic trapping studied by statistical polymer chain analysis*, J. Mol. Biol., **264**, 919-932, (1996).

[240] J. F. Marko and E. D. Siggia, *Statistical mechanics of supercoiled DNA*, Phys. Rev. E, **52**, 2912, (1995).

[241] J. P. Peters, L. S. Mogil, M. J. McCauley, M. C. Williams, and L. J. Maher, *Mechanical properties of base-modified DNA are not strictly determined by base stacking or electrostatic interactions*, Biophys. J., **107**, 448-459, (2014).

[242] M. Fernández-Sierra, Q. Shao, C. Fountain,L. Finzi, and D. Dunlap, *E. coli Gyrase Fails to Negatively Supercoil Diaminopurine-Substituted DNA*, J. Mol. Biology, **427**, 2305-2318, (2015).

[243] L. Bongini, V. Lombardi, and P. Bianco, *The transition mechanism of DNA overstretching: a microscopic view using molecular dynamics*, J. R. Soc. Interface, **11**, 20140399, (2014).

[244] S. B. Smith, Y. Cui, and C. Bustamante, *Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules*, Science, **271**, 795-799, (1996).

[245] X. Zhang, H. Chen, H. Fu, P. S. Doyle, and J. Yan, *Two distinct overstretched DNA structures revealed by single-molecule thermodynamics measurements*, Proc. Natl. Acad. Sci., **109**, 8103-8108, (2012).

[246] X. Zhang, Y. Qu, H. Chen, I. Rouzina, S. Zhang, P. S. Doyle, and J. Yan, *Interconversion between three overstretched DNA structures*, J. Am. Chem. Soc., **136**, 16073-16080, (2014).

[247] J. F. Allemand, D. Bensimon, R. Lavery, and V. Croquette, *Stretched and overwound DNA forms a Pauling-like structure with exposed bases*, Proc. Natl. Acad. Sci., **95**, 14152-14157, (1998).

[248] J. Kypr, I. Kejnovska, D. Renciuk and M. Vorlickova, *Circular dichroism and conformational polymorphism of DNA*, Nucleic Acids. Res., **37**, 1713-1725, (2009).

[249] J. P. Peters, S. P. Yelgaonkar, S. G. Srivatsan, Y. Tor, and L. James Maher, *Mechanical properties of DNA-like polymers*, Nucleic Acids Res., **41**, 10593-10604, (2013).

[250] S. Whitelam, S. Pronk, and P. L. Geissler, *There and (slowly) back again: Entropy-driven hysteresis in a model of DNA overstretching*, Biophys. J., **94**, 2452-2469, (2008).

[251] X. Zhang, H. Chen, S. Le, I. Rouzina, P. S. Doyle, and J. Yan, *Revealing the competition between peeled ssDNA, melting bubbles, and S-DNA during DNA overstretching by single-molecule calorimetry*, Proc. Natl. Acad. Sci., **110**, 3865-3870, (2013).

[252] J. W. J. Kerssemakers, E. L. Munteanu, L. Laan, T. L. Noetzel, M. E. Janson and M. Dogterom,*Assembly dynamics of microtubules at molecular resolution*, Nature, **442**, 709-712, (2006).

[253] T. Aggarwal1, D. Materassi, R. Davison, T. Hays, and M. Salapaka, *Detection of Steps in Single Molecule Data*, Cell Mol Bioeng., **5**(1), 14-31, (2012).

[254] J. M Eeftens, S. Bisht, J. Kerssemakers, M. Kschonsak, C. H. Haering, and C. Dekker, *Real-time detection of condensin-driven DNA compaction reveals a multistep binding mechanism*, The EMBO Journal, **37**(23), 3448-3457, (2017).

[255] A. Umbarger, E. Toro, M. A.Wright, G. J. Porreca, D. Baù, S.-H. Hong, M. J. Fero, L. J. Zhu, M. A. Marti-Renom, H. H. McAdams, L. Shapiro, J. Dekker, and G. M. Church. *The three dimensional architecture of a bacterial genome and its alteration by genetic perturbation.*, Molecular cell,**44**, 252–264, (2011)

[256] C. Haass, and D. J. Selkoe, *Cellular processing of beta-amyloid precursor protein and the genesis of amyloid beta-peptide*. Cell. **75**, 1039–1042, (1993).

[257] D. J. Selkoe, J. and Hardy, *The amyloid hypothesis of Alzheimer's disease at 25 years*. EMBO Mol. Med. **8**, 595–608, (2016).

[258] V. I. Zannis, A. Chroni, and M. Krieger. *Role of apoA-I, ABCA1, LCAT, and SR-BI in the biogenesis of HDL*. J. Mol. Med., **84**, 276–294,(2006).

[259] J. F. Oram, and J. W. Heinecke. *ATP-binding cassette transporter A1: a cell cholesterol exporter that protects against cardiovascular disease*, Physiol. Rev., **85**, 1343–1372, (2005).

[260] B. A. Kingwell, M. J. Chapman,A. Kontush, and N. E. Miller. *HDLtargeted therapies: progress, failures and future*. Nat. Rev. Drug Discov., **13**, 445–464, (2014).

[261] C. Vitali, C. L. Wellington, and L. Calabresi. *HDL and cholesterol handling in the brain*. Cardiovasc. Res., **103**, 405–413, (2014).

[262] M. Gregori,V. Cassina, D. Brogioli, D. Salerno, L. De Kimpe, W. Scheper, et al., *Stability of Ab (1-42) peptide fibrils as consequence of environmental modifications*. Eur. Biophys. J., **39**, 1613–1623, (2010).

[263] N. E. Persson, M. A. McBride, M. A. Grover, E. Reichmanis, *Automated Analysis of Orientational Order in Images of Fibrillar Materials*, Chem. Mater., **29**, 3−14, (2017).

[264] Z. P. Guven, P. H. Silva, Z Luo, U. B. Cendrowska, M. Gasbarri, S. T. Jones, and F. Stellacci, *Synthesis and Characterization of Amphiphilic Gold Nanoparticles.* J. Vis. Exp., **149**, e58872, (2019).

[265] R. Amit, A. B. Oppenheim, and J. Stavans, *Increased Bending Rigidity of Single DNA Molecules by H-NS, a Temperature and Osmolarity Sensor*, Biophys. J., **84**, 2467-2473, (2003).

[266] Y. Liu, H. Chen, L. J. Kenney, and J. Yan, *A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes*, Genes Dev., **24**, 339-344, (2010).

[267] C. J. Lim, L. J. Kenney, and J. Yan, *Single-molecule studies on the mechanical interplay between DNA supercoiling and H-NS DNA architectural properties*, Nucleic Acids Res., **42**, 8369-8378, (2014).

[268] R. S. Winardhi, S. Castang, S. L. Dove, J. Yan, *Single-Molecule Study on Histone-Like Nucleoid- Structuring Protein (H-NS) Paralogue in Pseudomonas aeruginosa: MvaU Bears DNA Organization Mode Similarities to MvaT*, PLoS One, **9**, e112246, (2014).

[269] F. C. Fang and S. Rimsky *New insights into transcriptional regulation by H-NS*, Curr. Opin. Microbiol., **11**, 113-120, (2008).

[270] A. S. Wegner, K. Wintraecken, R. Spurio, C. L. Woldringh, R. de Vries, T. Odijk, *Compaction of isolated Escherichia coli nucleoids: Polymer and H-NS protein synergetics.*, J Struct Biol., **194**(1), 129-37, (2016).

[271] A. Stylianou, D. Yova, *Surface nanoscale imaging of collagen thin films by Atomic Force Microscopy.* Mater. Sci. Eng. C. Mater. Biol. Appl., **33**, 2947–2957, (2013).

[272] A. Stylianou, *Atomic Force Microscopy for Collagen-Based Nanobiomaterials*. J. Nanomater., 1–14, (2017).

[273] K. Mizuno, E. Adachi, Y. Imamura, O. Katsumata, T. Hayashi, *The fibril structure of type V collagen triple-helical domain*. Micron, **32**, 317–323, (2001).

[274] Š. Rýglová, M. Braun, T. Suchý, *Collagen and Its Modifications—Crucial Aspects with Concern to Its Processing and Analysis*, Macromol. Mater. Eng., **302**, 1600460, (2017).

[275] J.K. Mouw, G. Ou, V.M. Weaver, *Extracellular matrix assembly: a multiscale deconstruction*. Nat. Rev. Mol. Cell Biol., **15**, 771–785, (2014).

[276] L. Puricelli, M. Galluzzi, C. Schulte, A. Podestá, and P. Milani, *Nanomechanical and topographical imaging of living cells by atomic force microscopy with colloidal probes.*, Review of Scientific Instruments, **86**, 033705, (2015).

[277] A. Cartagena and A. Raman, *Local Viscoelastic Properties of Live Cells Investigated Using Dynamic and Quasi-Static Atomic Force Microscopy Methods*, Biophysical Journal, **106**, 1033-1043, (2014).

[278] S. C. Lieber, N. Aubry, J. Pain, G. Diaz, S.-J. Kim, and S. F. Vatner, *Aging increases stiffness of cardiac myocytes measured by atomic force microscopy nanoindentation.*, Am. J. Physiol. Heart Circ. Physiol. **287**, H645 (2004).

[279] T. K. Berdyyeva, C. D.Woodworth, and I. Sokolov, *Human epithelial cells increase their rigidity with ageing in vitro: direct measurements.*, Phys. Med. Biol. **50**, 81, (2005).

[280] S. E. Cross, Y. S. Jin, J. Rao, and J. K. Gimzewski, *Nanomechanical analysis of cells from cancer patients.*, Nat. Nanotechnol., 2, 780, (2007).

[281] S. Iyer, R. M. Gaikwad,., I. Sokolov. *AFM detects differences in the surface brush on normal and cancerous cervical cells.*, Nat. Nanotechnol., **4**, 389–393, (2009).

[282] L. Bastatas, D. Martinez-Marin, ., S. Park. *AFM nano-mechanics and calcium dynamics of prostate cancer cells with distinct metastatic potential.*, Biochim. Biophys. Acta., **1820**, 1111–1120, (2012).

[283] L. Scarfò, A. J. Ferreri, P. Ghia, *Chronic lymphocytic leukaemia*, Crit Rev Oncol Hematol., **104**, 69-82, (2016).

[284] E. Hacken, J. A. Burger, *Microenvironment interactions and B-cell receptor signaling in Chronic Lymphocytic Leukemia: Implications for disease pathogenesis and treatment.*, Biochim. Biophys. Acta, **1863**, (3), 401-413, (2015).

[285] Y. Yamanashi, T. Fukuda, H. Nishizumi, T. Inazu, K. Higashi, D. Kitamura, T. Ishida, H. Yamamura, T. Watanabe, and T. Yamamoto, *Role of tyrosine phosphorylation of HS1 in B cell antigen receptor-mediated apoptosis.*, J Exp Med., **185**, (7), 1387-92, (1999).

[286] C. Scielzo, P. Ghia, A. Conti, A. Bachi, G. Guida, M. Geuna, M. Alessio, and F. Caligaris-Cappio, *HS1 protein is differentially expressed in CLL patient subsets with good or poor prognoses*, J Clin Invest., **115**, (6), 1644-5, (2005).

[287] E. ten Hacken, C. Scielzo, M. T. Bertilaccio, L. Scarfò, B. Apollonio, F. Barbaglio, K. Stamatopoulos, M. Ponzoni, P. Ghia, and F. Caligaris-Cappio, *Targeting the LYN/HS1 signaling axis in CLL*, Blood, **121**,(12), 2264-73, (2013).

[288] C. Scielzo, M. T. Bertilaccio, G. Simonetti, A. Dagklis, E. ten Hacken, C. Fazi, M. Muzio, V. Caiolfa, D. Kitamura, U. Restuccia, A. Bachi, M. Rocchi, M. Ponzoni, P. Ghia, F. Caligaris-Cappio. *HS1 has a central role in the trafficking and homing of leukemic B cells*, Blood, **116**,(18), 3537-46, (2009).

[289] S.R. Nielsen, V. Quaranta, A. Linford, P. Emeagi, C. Rainer, A. Santos, L. Ireland, T. Sakai, K. Sakai, Y.S. Kim, D. Engle, F. Campbell, D. Palmer, J.H. Ko, D. A. Tuveson, E. Hirsch, A. Mielgo, M.C. Schmid, *Macrophage-secreted granulin supports pancreatic cancer metastasis by inducing liver fibrosis.*, Nat Cell Biol., **18**(5), 549-60, (2016).

[290] A.V. Nguyen, K.D. Nyberg, M.B. Scott, A.M. Welsh, A.H. Nguyen, N. Wu, S.V. Hohlbauch, N.A. Geisse, E.A. Gibb, A.G. Robertson, T.R. Donahue, A.C. Rowat, *Stiffness of pancreatic cancer cells is associated with increased invasive potential.*, Integr Biol, **8**,(12), 1232-1245, (2016).

[291] J. Zhu, Q. Li, J.T. He, G.Y. Liu, *Expression of TAK1/TAB1 expression in non-small cell lung carcinoma and adjacent normal tissues and their clinical significance.*, Int J Clin Exp Pathol., **8** (12), 15801-7, (2015).

[292] M. Rubinstein, R. H. Colby *Polymer Physics*, OUP Oxford, (2003).

[293] C. Bustamante , J.F. Marko, E. D. Siggia , S. Smith, *Entropic elasticity of lambda-phage DNA*, Science, **265**(5178), 1599-1600, (1994).

[294] G. L.M. Fisher, C. L. Pastrana, V. A. Higman, A. Koh, J. A. Taylor, A. Butterer, T. Craggs, F. Sobott, H. Murray, M. P. Crump, F.o Moreno-Herrero and M. S. Dillingham , *The structural basis for dynamic DNA binding and bridging interactions which condense the bacterial centromere*, eLife, e28086, (2017).