Methods 197 (2022) 39–53

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Reconstruction of mechanical unfolding and refolding pathways of proteins with atomic force spectroscopy and computer simulations

Qing Li, Dimitra Apostolidou, Piotr E. Marszalek *

Department of Mechanical Engineering and Materials Science, Duke University, 27708 Durham, NC, United States

ARTICLEINFO

Keywords:
Atomic Force Microscopy
Single Molecule Force Spectroscopy
Steered molecular dynamics simulations
Protein unfolding
Repeat proteins
Coiled coils

ABSTRACT

Most proteins in proteomes are large, typically consist of more than one domain and are structurally complex. This often makes studying their mechanical unfolding pathways challenging. Proteins composed of tandem repeat domains are a subgroup of multi-domain proteins that, when stretched, display a saw-tooth pattern in their mechanical unfolding force extension profiles due to their repetitive structure. However, the assignment of force peaks to specific repeats undergoing mechanical unraveling is complicated because all repeats are similar and they interact with their neighbors and form a contiguous tertiary structure. Here, we describe in detail a combination of experimental and computational single-molecule force spectroscopy methods that proved useful for examining the mechanical unfolding and refolding pathways of ankyrin repeat proteins. Specifically, we explain and delineate the use of atomic force microscope-based single molecule force spectroscopy (SMFS) to record the mechanical unfolding behavior of ankyrin repeat proteins and capture their unusually strong refolding propensity that is responsible for generating impressive refolding force peaks. We also describe Coarse Grain Steered Molecular Dynamic (CG-SMD) simulations which complement the experimental observations and provide insights in understanding the unfolding and refolding of these proteins. In addition, we advocate the use of novel coiled-coils-based mechanical polypeptide probes which we developed to demonstrate the vectorial character of folding and refolding of these repeat proteins. The combination of AFM-based SMFS on native and CC-equipped proteins with CG-SMD simulations is powerful not only for ankyrin repeat polypeptides, but also for other repeat proteins and more generally to various multidomain, non-repetitive proteins with complex topologies.

1. Introduction

Single Molecule Force Spectroscopy (SMFS) comprises a set of techniques developed about 30 years ago to enable mechanical manipulations of *individual* macromolecules, primarily biopolymers such as DNA, proteins and polysaccharides in order to examine their structural and mechanical properties [1–23]. Regardless of the details of the technical platform used for SMFS measurements, individual macromolecules are attached between two objects which can be moved relative to one another with sub-nanometer precision so that the extension of the molecule can be accurately controlled. One of these two objects is either a force sensor directly responding to the molecular tension by undergoing elastic deformation, measured as a deflection from the equilibrium shape, or is trapped in a potential well so its position relative to the center of the well can be monitored and used to determine the force exerted on the probe. The first approach is exploited in atomic force microscopy (AFM) [24–28] based SMFS, where a micromachined thin cantilever beam terminated with a sharp tip is typically used as a force sensor and a glass (polystyrene) bead trapped by a focused laser beam is used in the second approach that is exemplified by laser tweezer [29–38] (optical trap)-based SMFS. Instead of a polystyrene bead, one can use a paramagnetic bead and expose it to a nonuniform magnetic field to produce a finely controlled pulling force, as is implemented in magnetic tweezers (MT) [39–45]-based SMFS. The results from SMFS measurements are typically expressed as a force versus extension relationship, when a molecule is stretched at a constant velocity or as length versus time relationship, when a molecule is subjected to a constant force or a force ramp (force clamp measurements) [46]. Over the years, SMFS was applied to directly measure the elasticity of double and single-stranded DNA and RNA in the absence or presence of various DNA enzymes [47–54], allowing to directly examine the applicability of polymer physics models such as the Freely Jointed Chain model (FJC) [42,51] or the Worm Like Chain model (WLC) [55] to nucleic acids and to probe

* Corresponding author.
E-mail address: pemar@duke.edu (P.E. Marszalek).

https://doi.org/10.1016/j.ymeth.2021.05.012
Received 17 March 2021; Received in revised form 14 May 2021; Accepted 15 May 2021

Available online 18 May 2021

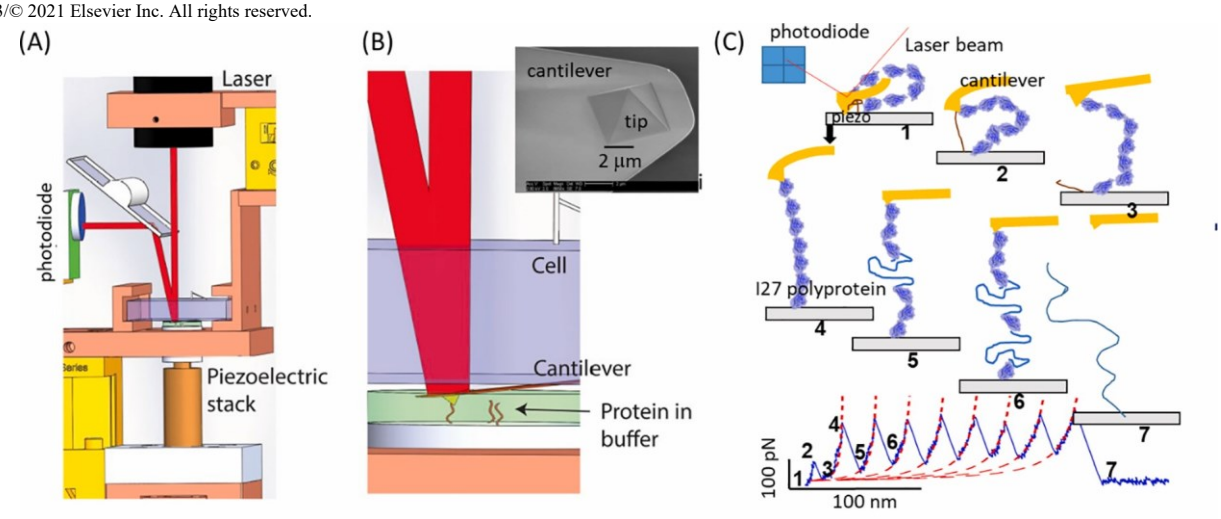


Fig. 1. (A) A simplified schematic of an AFM apparatus for SMFS (B) close-up of the fluid cell with the AFM cantilever mounted and a protein sample. Top right corner: SEM image of a cantilever with an integrated pyramidal tip. A laser beam is reflected off of the back of the cantilever and when projected onto a split photodiode, the cantilever's bending is measured by the displacement of the laser beam on the photodiode. The cantilever is mounted at the bottom of a quartz fluid cell and is immersed in a solution surrounding the sample of proteins adsorbed on a clean glass surface. A piezoelectric actuator (visible in (A) as piezoelectric stack) controls the Z position of the sample relative to the cantilever, allowing the cantilever's tip to probe the sample on approach and to pick up a protein molecule and stretch it as shown in (C) so its force-extension relationship (force spectrum) can be determined. A mechanical unfolding signature of a synthetic "polyprotein" composed of eight identical repeats of titin's I27 domain is shown at the bottom of (C). (C) Reproduced from our ref. [21].

the effect of molecular tension on the mechanoenzymatics of DNA and RNA polymerases, gyrases and topoisomerases [52,56–62]. Employing SMFS to various polysaccharide chains captured interesting force-induced conformational transitions within polysaccharide backbones and even within sugar rings [63–69]. SMFS found numerous applications to examine the elasticity of proteins with known and unknown mechanical functions [25,70–79] and to evaluate the strength of various inter- and intra- molecular interactions such as e.g. between receptors and their ligands or interactions involved in molecular recognition and adhesion [27,80–93]. In addition, SMFS has been broadly applied to study protein unfolding and refolding reactions, allowing new insights into (un)folding pathways [4,10,22,23,28,31,33–35,37,41,70,76,90,94–138].

2. Experimental SMFS of proteins by AFM

The early application of SMFS to examine the elasticity of muscle protein titin [139–141] revealed its great potential, not only for studying proteins with clear mechanical functions (such as titin), but generally to directly probe the structural stability and mechanical unfolding and refolding reactions of many proteins without known mechanical functions, providing novel insights into protein folding mechanisms and folding energy landscapes. Typically, in SMFS unfolding and refolding measurements, the external force is applied through the N and C termini of the protein defining a unique mechanical (un)folding reaction coordinate coinciding with the direction of the NC vector with its magnitude being equal to the distance between the termini. Thus, mechanical (un) folding measurements generally probe different unfolding and refolding pathways as compared to those pathways under chemical or thermal denaturation/renaturation conditions. By attaching pulling handles (polypeptide chains or DNA strands) to residues different than the termini (e.g. by exploiting native or engineered cysteine residues at specific locations), it is possible to mechanically unravel the protein along an (almost) arbitrary mechanical unfolding reaction coordinate, thus probing protein energy landscapes [95,142–147]. While unfolding and refolding protein force spectrograms reveal a wealth of information about mechanical pathways and can capture mechanical unfolding and refolding intermediates, it is generally rather difficult, particularly for large multidomain proteins with complex 3D structures, to identify the origin and order of unfolding events at the structural level. By carefully analyzing length increments following various partial unfolding events, it is possible to determine quite accurately the number of amino acids contributing to each event [28,146,148], which in conjunction with the detail structural information about the fold or some modeling studies (e.

g. molecular dynamics computer simulations) may help recreate (un) folding pathways. In addition, protein engineering that involves creating hybrid proteins or truncated variants are valuable in mapping the unfolding events onto the 3D protein structure, which in turn may locate key intramolecular interactions stabilizing the protein fold.

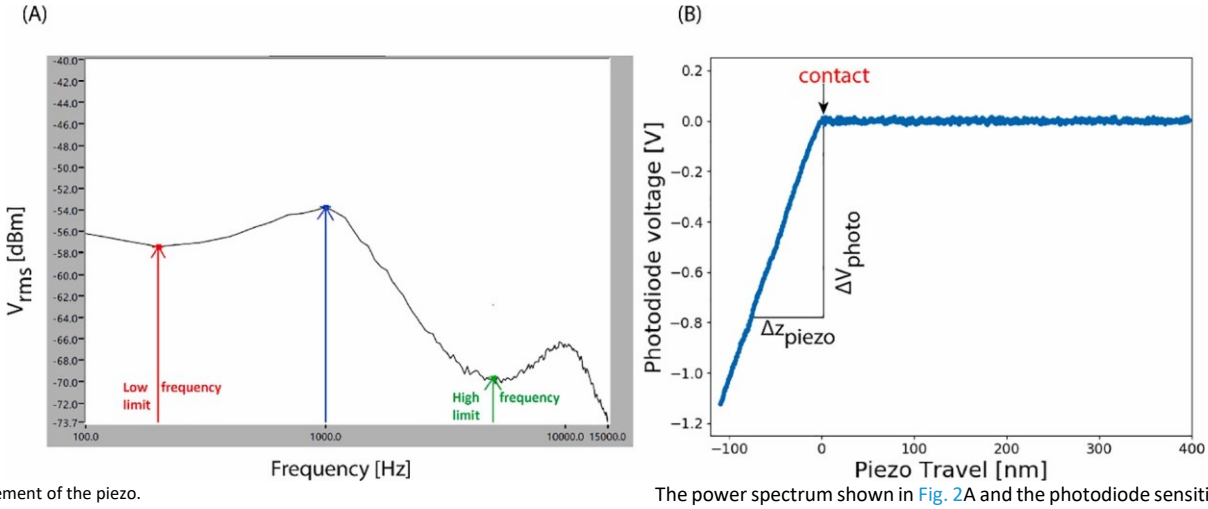
2.1. AFM force spectrometers

Many experimental details of SMFS on proteins, their attachment to AFM cantilevers and substrates as well as the description of the instrumentation involved can be found in various original papers and reviews, e.g. [3,7,98,149–154]. In Fig. 1A and B we show a simplified schematic of a SMFS measurement on a protein molecule using an atomic force microscope (AFM)-based force spectrometer. Any commercial AFM, which is equipped with a piezo position sensor at least in the pulling direction (typically, in the Z direction) may be used or adapted to be used for force spectroscopy measurements. Building an inexpensive force spectrometer "from scratch" (at a cost ~ 20,000 USD) is also possible, and various designs as well as the control software for data acquisition are available in the literature [149,155].

2.2. AFM cantilevers and their calibration

There are many types of commercially available AFM cantilevers that are appropriate for SMFS measurements on proteins using either commercial or home-made force spectrometers. V-shaped Silicon nitride cantilevers such as, e.g. MLCT cantilevers from Bruker or diving board- shaped Biolever cantilevers from Olympus (available e.g. from Bruker or Asylum Research/Oxford Instruments) are small and soft (spring constant in the range of 5–40 pN/nm) and sensitive enough to measure forces greater than approximately 5 pN (in the 1 kHz bandwidth), which allows many SMFS measurements even on relatively mechanically weak proteins that may unfold at forces around 20 pN (or greater), when stretched at the rate of 5–1000 nm/s. Because spring constants for commercially available cantilevers are given for orientation only, as

Fig. 2. (A) Power spectrum of an AFM cantilever measured in water. (B) Photodiode signal generated by deflecting the end of the cantilever by a known distance due to the movement of the piezo.



they do not accurately reflect the actual values which may vary greatly even in the same batch, it is absolutely necessary to calibrate each cantilever before SMFS measurements. The most commonly used method to determine the spring constant of AFM cantilevers, k_c (pN/nm = mN/m) is based on the energy equipartition theorem, which posits (when applied to cantilevers) that at thermal equilibrium, the average elastic potential energy of the cantilever is related to k_BT, where k_B is the Boltzmann constant and T is the absolute temperature, in K, at which SMFS measurements are carried out [156–158]. Applying the equipartition theorem to the cantilever deformations one obtains:

$$(½)k_C < z^2_{cantilever} > = (½)k_BT \quad (1)$$

At 298 K, k_BT = 4.11 pN nm, a value worth remembering. Thus, observing thermal motion of an AFM cantilever and determining $< z^2_{cantilever} >$ allows one to determine its k_C in a noninvasive fashion. This process actually involves two steps:

a) A power spectrum of a free cantilever undergoing random thermal motion (importantly, while being away from the substrate) is measured (Fig. 2A) so that the average squared amplitude of the photodiode voltage (generated by the laser beam reflected off of the top of the cantilever), $< V^2_{photodiode} >$, under the first peak (centered in Fig. 2A around 1 kHz, blue arrow) can be determined. Note, that the power spectrum may reveal some non-thermal movement of the cantilever (likely due to the mechanical noise in the environment) that manifests itself as a slight increase of the photodiode voltage at low frequencies (below ~ 200 Hz, in Fig. 2A, red arrow). This part of the spectrum should not be included in the $< V^2_{photodiode} >$ analysis as it does not belong in the energy equipartition due to the thermal agitation of the cantilever, and only the part between the red and green vertical lines (in Fig. 2A) should be considered (note that at frequencies greater than the frequency marked by green arrow, the second harmonic peak becomes visible).

b) The relationship between the magnitude of the photodiode voltage and cantilever bending needs to be established to convert $< V^2_{photodiode} >$ into $< z^2_{cantilever} >$. This is accomplished by producing static deformation of the cantilever in a controlled fashion. A *clean and rigid* substrate (e.g. a glass coverslip) is mounted on the piezoelectric actuator of the AFM, which moves it to contact the AFM cantilever tip (arrow in Fig. 2B) and presses it further by a known distance Δz_{piezo} while producing a photodiode voltage ΔV_{photo} (Fig. 2B). Since the tip is in contact with the piezo during this deformation it can be assumed that $\Delta z_{cantilever} = \Delta z_{piezo}$, establishing a relationship between the cantilever bending and the photodiode voltage. It is important that this step is performed not on a protein sample, but on a clean and rigid substrate to avoid possible deformations of the protein layer on the substrate, which would then result in $\Delta z_{cantilever} < \Delta z_{piezo}$. With a known k_c, the force acting on the cantilever can be easily determined as $\boldsymbol{F = k_c \Delta z_{cantilever}}$.

The power spectrum shown in Fig. 2A and the photodiode sensitivity graph (Fig. 2B) were determined for a MLCT type "C" cantilever (Bruker). The sensitivity of the photodiode for this cantilever was determined to be 94.3 nm/V, and the spring constant was found to be 14.8 pN/nm.

Note that many AFM cantilevers are coated on the back (or on both sides) with a thin layer of gold to increase their reflectivity. The presence of gold, however, seems to increase their long-term drift in solution and it has been recommended by some researchers to remove this layer of gold which may significantly improve cantilever's stability [7,159]. Ideally, one's AFM is equipped with an infrared laser for this approach to work well, as a red laser light is weakly reflected from silicon nitride which results in a weak photodiode signal.

2.3. Substrates for SMFS of proteins by AFM

2.3.1. Simple glass slides/disks

For the calibration of the cantilever, clean round glasses, typically of 15 mm diameter, are used (Ted Pella, Catalogue number 26021). These glass slides are usually rinsed well with ethanol and air dried prior to placing them on a magnetic holder (Ted Pella, Catalogue Number 16218) of the same diameter by using a double-sided adhesive tab (Ted Pella, 16079). Clean glass slides can also be used to deposit samples for SMFS measurements, even though gold-coated glass slides are more commonly used.

2.3.2. Gold-coated glass disks

The same glass slides used for calibration of the cantilever are also used to deposit the sample in solution for SMFS measurements. Initially, the glass slides are cleaned by leaving them overnight in Piranha solution, which is an extremely corrosive solution and was handled with extra care. Then, they are carefully rinsed with water and air dried. The clean glass slides are then taken to an E-Beam Metal Evaporator (CHA Industries, Inc.) to coat 80 nm of Cr followed by 250 nm of Au. The Cr layer is necessary in allowing Au adhesion to glass disks. *2.4. Attachments of proteins for SMFS by AFM*

2.4.1. Nonspecific physical adsorption (physisorption)

Almost all proteins adsorb strongly to glass, gold, silicon, and silicon nitride surfaces and this physical adsorption can be exploited to perform SMFS measurements by AFM [151]. Simply, a drop of a solution containing a given protein at concentration of 10–1000 μg/ml is deposited on a clean surface (glass, gold-coated glass), allowing the protein molecules to adsorb to the surface. This step typically takes 10–30 min.

Samples should be protected from drying, as dehydration will denature many proteins. After that, the sample is gently washed a few times to remove the unbound protein. This is typically done by pipetting in and out a few tens of microliters of the buffer solution without bubbling. At this point, the sample can be placed in an AFM instrument. When contact is made between the cantilever tip and the sample of adsorbed proteins, the protein molecules will

also adsorb randomly to the tip surface, so upon the cantilever's withdrawal, a molecular bridge may be formed (or not) between the glass surface and the cantilever tip as shown in Fig. 1C. By adjusting the protein concentration in the sample before deposition for the adsorption and checking various contact forces (100–1000 pN) and contact times between the sample and the cantilever, it is possible to establish the "right" conditions to pick up one or a more molecules for SMFS measurements by multiple trials. A protein bridge is formed when upon the withdrawal of the cantilever from the sample surface a force acting on the cantilever causes it to bend towards the sample and be captured (Fig. 1C step 2). This is the easiest method for attaching proteins to surfaces and AFM tips for SMFS measurements. However, this method has several drawbacks. *First*, the parts of the protein that stick to glass/gold and the tip are likely denatured on contact so that only some proteins, such as polyproteins composed of tandem repeats of the same domain (module) may be measured, as the middle domains that are not denatured are representative of the whole protein, while the terminal domains are sacrificed as pulling "handles". A variation of this method involves flanking a given protein to be studied with polyproteins [115] that protect the central protein which are used as pulling handles. Polyproteins also provide a unique signature for single molecule measurements (e.g., a saw-tooth pattern, as in Fig. 1C**)** allowing to easily distinguish those measurements from undesired measurements on multiple molecules. However, adding polyprotein handles also significantly increases the molecular weight of the construct, which may complicate protein expression and purification. *Second and most important*, the success rate for picking up a single molecule is very low, typically less than 1/1000. Thus, hundreds of thousands of attempts need to be made to collect enough force-extension data. This can be achieved by automating AFM operation to carry out tens of thousands of force-extension measurements in an unsupervised fashion [150]. We note that this very simple, but inefficient method was used in the studies described later in this article. However, the reader is strongly encouraged to employ new, more complex attachment methods that are briefly outlined below, as these significantly increase the success rate of SMFS measurements on proteins (sometimes >1/10 attempts, as demonstrated e.g. by the Perkins group [160]).

2.4.2. Specific attachment of proteins to surfaces and AFM tips

Several groups pursued the development of various controlled attachment methods of proteins for efficient SMFS measurements, but it is beyond the scope of this article to describe many of the excellent approaches developed and tested, so we mention only briefly some of the more recent assays. For a more comprehensive review of different attachment methods, the reader is referred to various books on the subject and review articles, e.g. [161–163], and to the original publications cited therein, for methodological details.

a) New streptavidin–biotin assays

The very strong binding between the tetrameric protein streptavidin and a small molecule vitamin H (biotin) found numerous applications in the field of biology and biotechnology and has been successfully employed in AFM-based force spectroscopy studies from their beginning (see Refs. [80,82]). Recently, the Perkins group and collaborators developed an efficient method to couple biotin to proteins via converting a terminal cysteine to formylglycine, to which biotin can be coupled via a HIPS-based reagent [160]. This biotin-terminated protein will bind to AFM tips functionalized with streptavidin. Another approach for the protein biotinylation involves adding to the protein, at the DNA level, an AviTag sequence to which biotin can be enzymatically attached using the biotin ligase BirA enzyme [41].

b) Cysteines

It is well known that cysteine residues bind to gold by forming a relatively strong Au-S bonds. This coupling has been exploited in SMFS of proteins that can be engineered to have one or two terminal cysteines. Gold coated AFM tips are available (e.g., OBL biolevers), but it is also easy to produce a thin gold film on AFM tips or glass substrates pre- coated with a chromium/nickel layer. The formation of gold film is typically achieved by evaporating gold under a high vacuum. It is advisable to rotate AFM cantilevers during gold coating to produce a uniform deposition of gold which will prevent permanent cantilever bending. Exploiting Cys-S-Au bonds, one can directly attach proteins to a gold-coated substrate or use gold coated AFM tips to pick up cysteine- terminated proteins that were immobilized on a substrate using another approach (for a directional binding) [164]. Proteins terminated with cysteine residues can also be attached to PEG-Maleimide functionalized glass surfaces through a (maleimide)-C-S-(cysteine) bond [162] or by using more complex approaches, such as by making use of PEG-Azide functionalized surfaces which exploit cysteine-maleimide –DBCO-Azide reaction [160].

c) Chloroalkane ligand – HaloTag

HaloTag is a mutated (33 kDa) protein derived from a bacterial Haloalkane Dehalogenase enzyme that was engineered to covalently bind a synthetic ligand, e.g. a chloroalkane ligand such as HaloTag Thiol O4 ligand (Promega), which in turn can be permanently attached to a surface using a PEG-Maleimide crosslinker (Thermo Scientific [165]) . By genetically fusing the protein of interest to the HaloTag domain, the protein can be attached to the functionalized surface through the HaloTag-HaloTag ligand bond for SMFS measurements. This approach worked very well for AFM- and magnetic tweezers-based SMFS for a variety of proteins [41,165].

d) (ybbR-CoA)/SFP

The 11 amino acids-long ybbR tag is recognized by an enzyme 4′-phosphopantetheinyl transferase from Bacillus subtilis (Sfp) that ligates this tag to coenzyme A (CoA). To immobilize a given protein for SMFS measurements, the ybbR tag must be genetically fused to the protein and Sfp ligates the tag-terminated protein to CoA. CoA that carries a thiol group is itself immobilized onto a glass surface functionalized with PEG- Maleimide. This approach also allows for the attaching of ybbR- equipped proteins to AFM tips when those are functionalized with PEG-Maleimide-CoA [162,166].

e) Dockerin-Cohesin

Dockerins and cohesins are protein modules exploited in cellulose degrading enzymatic machineries, where they provide exceptionally strong interactions enabling dockerin-terminated enzymes to attach to the cohesins-terminated scaffoldin assembly [167]. The mechanical strength of the dockerin-cohesin adhesion complex was determined to be>500 pN even at relatively low loading rates of 2–5 nN/s [168,169]. By terminating the protein of interest with the dockerin tag and attaching cohesin to the AFM tip, it is then possible to specifically pick up the protein of interest for SMFS measurements, which can be performed on strong proteins with the mechanical stability of up to 500 pN [166]. The dockerin-cohesin system can be used alone or in combination with the ybbR-CoA attachment method, where the former method is used to attach the protein of interest to the AFM tip and the latter method can be used to immobilize the protein of interest on a surface [166].

f) Asparaginyl endopeptidase

Recently, an engineered mutant (C247A) of the enzyme, asparaginyl endopeptidase from the plant *oldenlandia affinis* (OaAEP1) [170] was shown to efficiently ligate the N and C termini of peptides and even folded proteins that are equipped with the N-terminal glycine-leucine (GL) sequence and the C-terminal asparagine-glycine-leucine (NGL) sequence [170]. This ligation reaction which occurs at very mild conditions was exploited for protein surface immobilization [171] and to engineer polyproteins [170] for SMFS measurements [172].

It is important to note that in practice, various methods of protein attachment are combined together depending on the specifics of the protein examined and the SMFS experiments planned such that the protein may be connected to a surface using e.g. a covalent attachment method (e.g. HaloTag, ybbR-CoA or SpyTag [173]) and can be picked up by the AFM tip exploiting e.g. a noncovalent biotin-streptavidin or dockerin-cohesin coupling [160,166] or thiol coupling to gold-coated tips [165].

3. Protein engineering for SMFS

The common practice in expressing proteins is using plasmids (vectors) which can be easily modified by the deletion or insertion of gene sequences.

This can be done by implementing various techniques, such as using restriction sites. The plasmids are then transformed in bacterial cells, which were modified to multiply the plasmids very effectively without the introduction of mutations, and the plasmids are then collected using a well-established process called miniprep. The sequence of the plasmids is verified after this step. A second transformation is performed in bacterial cells specialized in the overexpression of proteins. Various proteins require different protocols, such as incubation times and temperatures, to optimize their expression, which is then followed by the collection of the protein and its purification. The purification step usually implements a purification tag, which is part of the protein construct. There are various tags available, with the most common one being the HisTag. After purification, the protein is then ready to be used for SMFS experiments. *3.1. Protein design, cloning, and expression*

Our NI6C-I27 construct consisted of six I27 domains in total which flanked the NI6C protein on both termini, (I27)3-NI6C-(I27)3. This chimeric polyprotein had its NI6C adapted from [174] with regards to the sequences of the capping and the internal repeats. The DNA sequence of NI6C was synthesized by Genescript (Piscataway, NJ). A Poly-I27 pRSETa vector was used to insert the NI6C gene (a kind gift from Dr. Jane Clarke, Cambridge University [175]) using restriction sites. The engineered plasmids were transformed into E. coli C41 (DE3) (Lucigen, Catalogue Number 89027), and the NI6C-I27 protein was expressed for 10 h using IPTG induction. The harvested cells were collected, lysed and purified by using a nickel affinity column (GE Healthcare, Catalogue Number 17–5268-02) followed by size exclusion HPLC.

3.2. Engineering CC probes and consensus ankyrin repeat hybrid proteins [23]

Codon optimization, synthesis and cloning into a pUC57 vector of gene sequences for CC, NI10C, NI4CCI6C, and NI8CCI2C was performed by Genescript. (I27)7-Streptag-pRsetA plasmid, adapted from Poly(I27)- pRsetA vector with the 8th I27 domain of the gene replaced by a Strep- tag, was also digested with the 4th I27 domain replaced by CC, NI10C, NI4CCI6C, or NI8CCI2C to create the designed constructs for SMFS measurements. The sequence of the four new plasmids was verified prior to protein expression. C41 (DE3) pLysS cells from Lucigen (Middleton, WI, USA) were used to overexpress (I27)3-CC-(I27)3-Streptag, (I27)3- NI10C-(I27)3-Streptag, (I27)3-NI4CCI6C-(I27)3-Streptag and (I27)3- NI8CCI2C-(I27)3-Streptag. Cells were grown for 4 h at 37 °C (until OD 600 > 0.8), followed by the addition of 0.2 mM IPTG and decrease of the temperature to room temperature for overnight expression. Cells were harvested the next day by spinning down at 4000 × g for 40 min and then frozen at − 80 °C for several hours. The cells were thawed and lysed and the supernatants were purified by gravity flow Strep-Tactin Sepharose column from IBA GmbH (Gottingen, Germany). ¨

4. Computer modeling of SMFS measurements on proteins

4.1. All-atom Steered molecular dynamics (SMD) simulations of proteins

Molecular dynamics simulations of SMFS experiments proved extremely helpful and almost indispensable for interpreting force- extension relationships and for gaining additional insights into the behavior of mechanical and non-mechanical proteins subjected to stretching forces [10,26,89,128,147,176–188]. Typically, to simulate a SMFS experiment, the protein of interest is immobilized at one end while a “spring” representing an AFM cantilever, and modeled with a harmonic potential, is attached to another terminus. The spring end is then moved at a constant velocity in the direction of the applied force (constant force simulations are also possible and those are designed to mimic “force clamp” SMFS experiments). These types of computer experiments, where an additional harmonic potential is added to a selected group of atoms within a protein whose total energy is modeled with one of the standard force-fields (e.g. AMBER or CHARMM) were coined steered molecular dynamics simulations (SMD) [189–191]. There are two major classes of SMD simulations of SMFS events. In the first approach termed “all atom simulations”, the protein itself, as well as the surrounding water environment, is modeled at a very high detail with all atoms (including hydrogen atoms) explicitly included in the model and subjected to MD calculations. Since a completely stretched protein is much longer as compared to the size of its native structure (e.g., a small I27 domain of titin composed of 89 amino acids becomes 32 nm long when fully stretched, compared to only 4 nm when in the natively folded state), to accommodate a completely stretched protein a very long water box needs to be used in all atom SMD simulations and the size of the whole system to be modeled reaches easily hundreds of thousands to tens of millions of atoms. Such systems are best modeled using supercomputers and even then, the speed at which the protein is stretched in the computer is orders (3–6 orders) of magnitude greater than in real AFM experiments. This potentially creates a problem, as forced protein unfolding is affected by the stretching speed (more accurately by the loading rate, in pN/s) and unfolding forces increase in magnitude with increasing loading rates [192,193]. This problem can be alleviated by performing simulations at decreasing loading rates and extrapolating the results to the ranges of experimental rates. Notwithstanding this caveat, all-atom SMD simulations can provide a plethora of atomistic details related to the response of protein’s structure to applied forces that cannot be deduced from experiments alone. In addition, only these types of simulations allow for the modeling of the effect of small structural variations within proteins, e.g. due to mutations, on the mechanical stability and force unfolding behaviors [179].

The second major class of SMD simulations involves a significant reduction of the complexity of all atom protein representations and an elimination of the solvent altogether, reducing the size of the system by 2–3 orders of magnitude. These so called “coarse-grained” SMD simulations are used in the remaining part of this article and their premise is described in some detail below. The main advantages of CG-SMD simulations of SMFS experiments on proteins are that stretching speeds *in silico* are significantly closer to the experimental speeds as compared to all-atom SMD simulations and that numerous simulations can be run in parallel even on a fairly simple computer system (see below) allowing to achieve better statistics. In addition, CG MD simulations uniquely allow to investigate refolding trajectories from unraveled (denatured) to native structures, as the energy function is based on the original protein structure and proteins refold to their initial states in a reasonable amount of computational time [124]. Among some disadvantages of this approach is the lack of atomistic details during mechanical unfolding and refolding, such as breaking and reforming of specific hydrogen bonds. Thus, it is more difficult to propose further experiments aimed at verifying the mechanical roles of certain residues through mutational approaches as compared to the results obtained with all atom simulations.

4.2. SMD simulations of unfolding and refolding of proteins using coarse-grained (CG) models

For our SMD simulations of AFM pulling measurements of proteins, we adapted a CG model described in Clementi et al [194]. In this particular model, each residue is represented by one pseudo atom centered at the location of the C_α atom of that amino acid. The interaction between non-neighboring pseudo atoms is based on native and non-native contacts determined for corresponding amino acids from high-resolution structural data of the protein (e.g. based on X-ray or NMR-determined structures). The energy function used in this CG model is given below:

$$E(\Gamma,\Gamma_0) = \sum_{bonds} K_r(r-r_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 + \sum_{dihedral} K_\varphi^{(n)}[1+\cos(n\times(\varphi-\varphi_0))] + \sum_{i<j-3}^{N} \left\{ \varepsilon(i,j)\left[5\left(\frac{r_{0,ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{0,ij}}{r_{ij}}\right)^{10}\right] + \varepsilon_2(i,j)\left(\frac{r_{0,ij}}{r_{ij}}\right)^{12}\right\} \quad (2)$$

The first three terms represent the energy associated with the deformation of bonds, bond angles and dihedral angles from their equilibrium

(native) positions. The term with ε(i, j) represents the non- local interactions based on the native structure and the term with ε_2(i, j) describes the energy penalty when non-native contacts are made. The subscript "0" is used to refer to the original native structure (X-ray or, NMR) while Γ denotes the current structure. K_r is the spring constant for all bonds between two pseudo atoms (Cα), with r being the distance between two consecutive Cα atoms; K_ϑ is the spring constant for the bond angles, where ϑ is the angle formed by three consecutive Cα pseudo atoms; (n) K_φ is the harmonic constant for dihedral angles with multiplicity of n; ϕ is the dihedral angle formed by four consecutive Cα atoms; $\varepsilon(i, j)$ and $\varepsilon_2(i, j)$ are the energetic parameters for native and non-native contacts, respectively; and r_{ij} is the distance between amino acids i and j. Using this energy function (force field) one can execute the SMD protocol using e.g. the GROMACS molecular dynamics package [195].

4.3. Computer resources for CG-SMD simulations

To carry out CG-SMD simulations even on fairly large proteins (e.g. exceeding 500 residues) one needs modest computer resources. A small Linux cluster with 8 to 64 computing cores and 16 to 64 GB of RAM, or a desktop computer with a multicore CPU and a GPU should be sufficient to carry out numerous CG-SMD simulations (to assure the convergence of the results) in a reasonable amount of time. The information provided below is for orientation purposes only, as benchmarks will vary depending on a specific hardware and software combination used. For example, when modelling a protein with 500 amino acids, the protein in a CG representation will be composed of 500 pseudo atoms only. On a Linux cluster composed of 4 AMD Opteron(TM) Processor 6274, with each CPU equipped with 16 computational cores, running at 2.2 GHz, one can run 8 CG-SMD simulations in parallel, using 8 cores for each simulation, as increasing the number of cores for such a small system (500 pseudo atoms) does not significantly increase the computational speed. Using the GROMACS 2018.2 package, one can then run a single 100 ns simulation in approximately one day, so one hundred of such CG- SMD simulations (using all 64 cores) can be performed in less than two weeks on a small computer system.

5. Reconstruction of vectorial unfolding and refolding pathways of repeat proteins by AFM and Coarse-Grained SMD simulations

In this section, we will illustrate the application of the above- described experimental and computational procedures to examine unfolding and refolding pathways of our model proteins composed of a 8 and 12 consecutive ankyrin repeat domains [22,23]. Ankyrin repeat

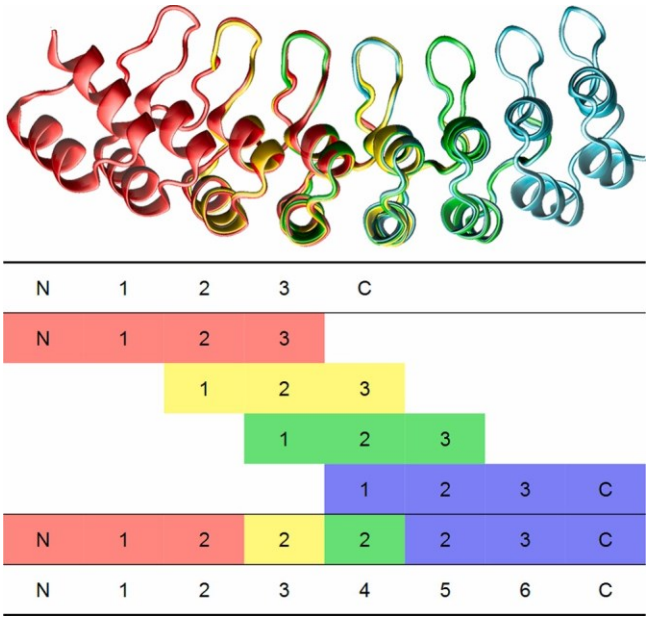


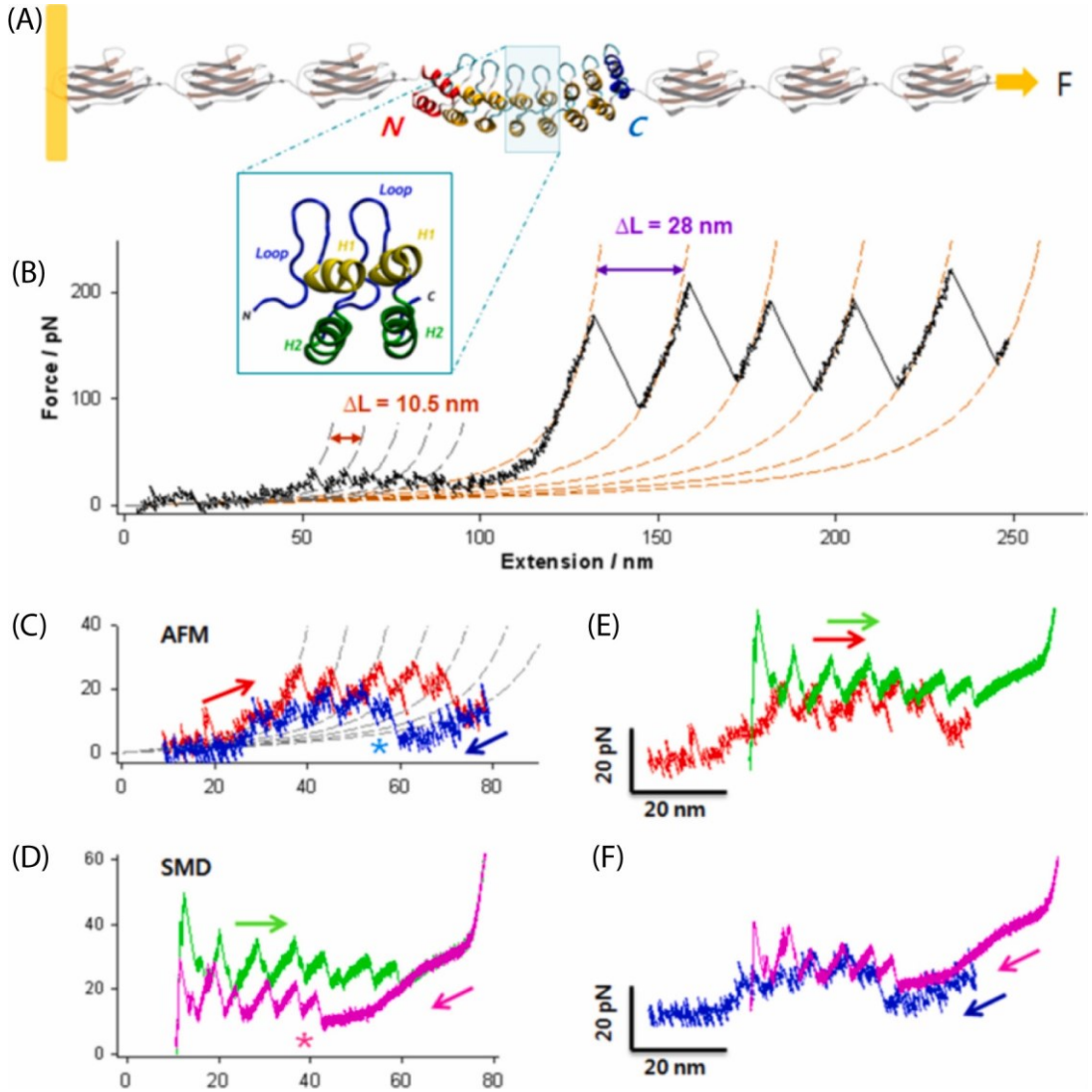
Fig. 3. RMSD fitting procedure 1. PDB structure is duplicated with 4 copies: red, yellow, green, and blue. 2. Internal repeats were aligned as follows: (2,3) Red with (1,2) Yellow, (2,3) Yellow with (1,2) Green, and (2,3) Green with (1,2) Blue 3. Final product of (N,1,2) Red-(2) Yellow-(2) Green-(2,3,C) Blue was recorded. 4. Both ends of the protein have coils composed by six glycine residues as the handles added to them. 5. Residue numbers from 1 to 265 (from N to C) were re-ordered to complete the structure of NI6C. Reproduced from our ref. [22] (and Supporting Information).

domains have been found in over 4700 proteins with diverse functions such as transcriptional initiators, cell cycle regulators, cytoskeletal, ion transporters, and signal transducers [196,197]. The ankyrin repeat motif is composed of 33 amino acids arranged into two alpha helices separated by loops. These folded peptides stack closely on top of each other to form extended solenoid structures (Fig. 3). By examining the sequences of many ankyrin repeats from diverse proteins, a model synthetic ankyrin repeat domain with the so-called consensus sequence was designed [198]. Synthetic proteins composed of various numbers of these consensus repeats were created with the smallest stable structure being NI3C composed of three consensus internal repeats (I) flanked by slightly different capping repeats N and C [198]. NI3C was crystalized and its atomic structure resolved (PDB code: 2QYJ) [174]. Consensus ankyrin repeat proteins proved to be extremely stable thermodynamically, even under the harshest conditions such as at temperatures up to 100˚C or under high concentrations of denaturants [174], making SMFS the only practical experimental approach to unfold the protein and examine its refolding properties.

5.1. AFM-based single molecule force spectroscopy

Purified NI6C-I27 protein was dialyzed in 20 mM Tris (pH 8.0), 150 mM NaCl, and 2 mM TCEP (Thermo Scientific, Catalogue Number 77720). 50 µl of a diluted protein solution (1–5 µg/ml) was incubated on a substrate for 20 min. The sample was then rinsed once by the addition and removal of 50 µl of the buffer, therefore removing protein molecules that did not get immobilized to the substrate surface. Both clean glass and nickel-NTA functionalized glass substrates were used for these AFM experiments. Both of these substrates gave similar force-extension curves of NI6C-I27. Custom-built AFM instruments equipped with an AFM detector head (Veeco Metrology Group (Now Bruker)), and high- resolution piezoelectric stages with vertical resolution of 0.1 nm (Physik Instrumente) were used to perform all stretching measurements. The spring constant, k_c, of each cantilever was calibrated, as described

Fig. 4. Complete mechanical unfolding and refolding traces of the (I27)3-NI6C-(I27)3 chimeric protein. (A) Depiction of the (I27)3-NI6C-(I27)3 construct. Moving from left to right, we have the gold surface followed by three I27 domains, NI6C, three I27 domains and the cantilever. NI6C contains eight ARs, with two internal repeats showing in the blue box, demonstrating the two helices (H1 and H2) and a loop, (B) Representative unfolding trace of NI6C-I27 with stretching speed of 100 nm/s. Two different WLC curves were used to fit the AFM data. For AR, we used a contour length increment of ΔLc=10.5 nm and persistence length p=0.78 nm (gray dashed lines), and for I27 domains we used ΔLc = 28 nm and p=0.36 nm (orange dashed lines). These values of ΔL are consistent with the fully stretched lengths of one AR and one I27 domain, respectively, (C) Unfolding (red) and refolding (blue) force extension traces of NI6C at 30 nm/s, with WLC fits (gray dashed lines) with ΔLc=10.5 nm and p=0.86 nm. Asterisk highlights the first refolding force peak appearing after the molecule was partially relaxed. (D) Simulated unfolding (green) and refolding (pink) force extension traces of NI6C. Both results show the first refolding force peak occurs only after the molecule has been significantly relaxed noted with an asterisk. (E) Overlapping of the unfolding force extension traces by SMD (green) and AFM (red). The shift in the SMD trace (to the right by 20 nm) compensates for the initial length of I27 domains that contribute to the extension in the AFM measurements but are absent in the SMD simulations. (F), comparison of the refolding force extension traces by SMD (pink) and AFM (blue) following the unfolding of NI6C. The same 20-nm shift was applied to the SMD trace. Reproduced from our reference [22].



previously, in solution using the energy equipartition theorem. An AFM cantilever tip was used to pick up protein molecules by gently touching the substrate. All force-extension measurements were performed at room temperature using Biolever cantilevers (OBL from Veeco (now Bruker), kc ≈ 6 pN/nm) at pulling speeds between 5 and 100 nm/s. The force peaks in the force-extension curves were fitted to the Worm-like- chain (WLC) model [55].

5.2. CG-SMD simulations of consensus ankyrin repeat proteins

NI6C was the first consensus AR protein that we studied by CG-SMD and contains an N-terminal, followed by 6 internal ankyrin repeats, and a C-terminal, and is composed of 253 amino acids. In the AFM experiments, I27 domains were used to flank the AR protein. Only 6 additional glycine residues were used, as the flanking sequence on each side in order to reduce the system size and focus the modeling on NI6C. In order to perform CG-SMD on NI6C, as shown in Fig. 3, NI3C (PDB: 2QYJ) was used to extrapolate NI6C. NI3C is the largest consensus AR protein with determined crystal structure. Building an accurate initial structure was critical given that our CG model employed a structure- based energy function. We used a root mean square displacement (RMSD) fitting procedure to build the structure of NI6C based on the X- ray crystallography of NI3C. In the fitting procedure, NI3C protein was divided into 5 domains (as demonstrated in Fig. 3) and fitting was performed to minimize the RMSD of the backbone atoms (Cα, C, O, N) among repeats 1, 2, and 3. Fig. 3 demonstrates a detailed step-by-step

Fig. 5. CG-SMD simulations of NI6C. (**A**) Unfolding; (**B**) Refolding. Note a folding nucleation event in the refolding simulation at t= 477 ns. Adapted from our ref. [22].

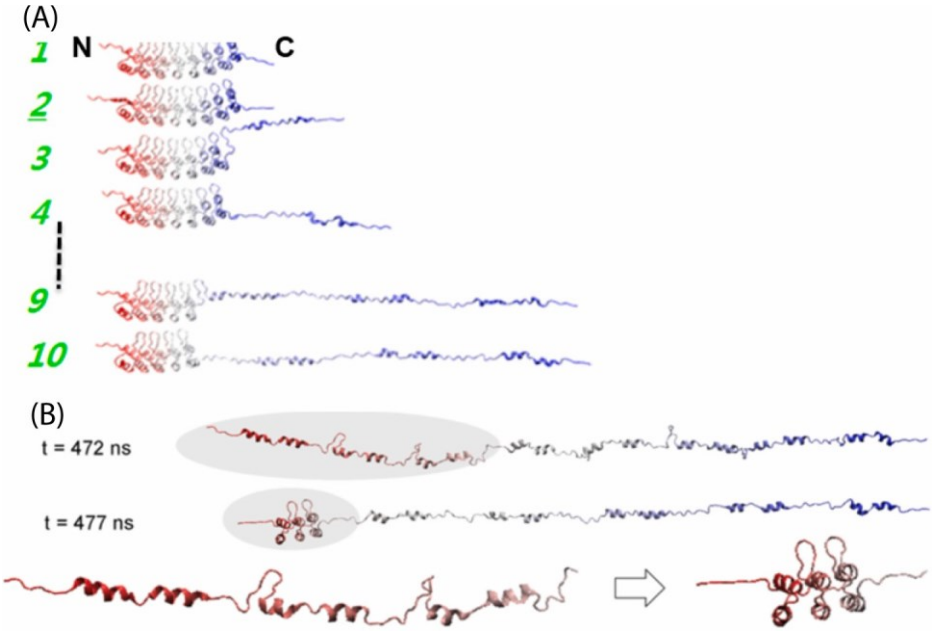


Fig. 6. (A) Native protein with a prominent anti-parallel CC structure (marked in red, PDB: 1NT2) adapted as our mechanical unfolding probe. (B) construct to examine the mechanical unfolding fingerprint of the CC probe. (C) Superposition of force extension curves obtained on the (I27)3-CC-(I27)3 construct. Adapted from

description of the procedure where the final structure built through the RMSD fitting is also shown with overlaps in different colors.

5.3. Combining AFM and computational SMFS experiments to reconstruct consensus ankyrin repeat protein unfolding and refolding pathways

A typical force extension relationship obtained by AFM stretching experiments on the (I27)3-NI6C-(I27)3 construct is shown in Fig. 4B. This measurement was obtained using an OBL biolever cantilever with a spring constant of 6 pN/nm, at a stretching speed of 100 nm/s.

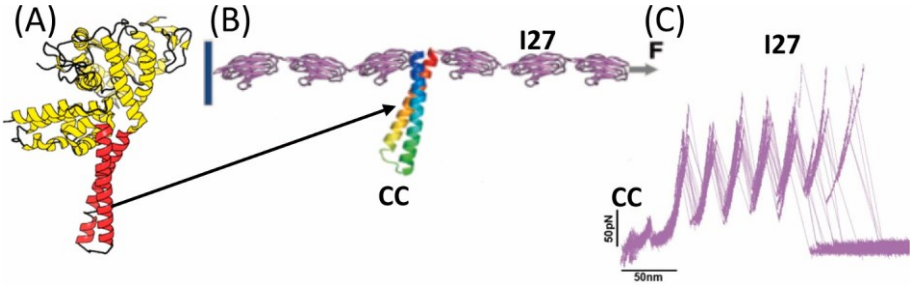
1. At extensions greater than ~ 100 nm, the recording reveals a pronounced and characteristic saw tooth pattern of large regular unfolding force peaks around 170–200 pN.
2. The experimental points at *(x, F)* coordinates along the rising phases of the recording leading to each large force peak can be reasonably fitted with the Worm-Like-Chain (WLC) model of polymer elasticity, which is frequently used to describe the elasticity of double stranded DNA, unstructured polypeptides and polyproteins composed of individually folded domains (such as polyI27) (dashed orange lines in Fig. 4B):

$F_{WLC} = (k_BT/L_P)\,[0.25\,(1 - x/L_C)^{-2} + x/L_C - 0.25\,]$ (3), where L_P is the so called persistence length and L_C is the contour length of the polymer [55].

3. The consecutive WLC fits were obtained with the same persistence length of 0.36 nm [141] and increasing contour length $L_C + \Delta L_C$, with ΔL_C determined to be 28 nm for all peaks [115].
4. Gathering the information collected from 1 through 3, one concludes that the large force peaks represent the mechanical unfolding fingerprint of the I27 domains flanking the NI6C insert. The number of I27 peaks is five, which is greater than the number of I27 domains on one side of the NI6C insert (which is three), indicating that the recording was obtained on the protein fragment that included at least five I27 domains, which in turn indicates that the stretched fragment must have included the complete NI6C insert as two out of the five domains must be located at the other flank. Determining the number of unfolding peaks corresponding to the fingerprinting domains (that also serve as pulling handles) is commonly used in SMFS (particularly when exploiting nonspecific protein attachment that results in random fragments to be stretched) to determine which recordings were performed on complete, full length proteins of interest (insert between the handles).
5. To identify the mechanical unfolding fingerprint of NI6C, we turn our attention to the part of the force curve at extensions < than 100 nm.

At these extensions, one can discern a set of much smaller force peaks at approximately 25 pN each which can be fitted with a family of WLC curves generated with L_C = 0.86 nm and ΔL_C = 10.5 nm (gray dashed lines in Fig. 4B, C.). This value of the persistence length is typical for alpha helical peptide bundles, such as e.g. spectrin [199] and suggests that these small peaks may represent the mechanical unfolding of ankyrin repeats. Each consensus ankyrin repeat contains 33 amino acids (a.a.) and in the fully stretched state is expected to have the length of 33 a.a. × 0.365 nm/a.a. [146] corresponding to 12 nm. Subtracting the folded length of each AR (around 1 nm) one obtains the expected ΔL_C of 11 nm, which agrees very well with the 10.5 nm determined from the WLC fits. Thus, one concludes that the small force peaks indeed correspond to the unfolding of NI6C.

6. It seems somewhat surprising and counterintuitive that the consensus ankyrin proteins, being thermodynamically very stable [174], unfold at such low forces. However, this observation supports the notion that unfolding pathways induced mechanically do not need to coincide with pathways induced by denaturants or by an increased temperature.
7. When one limits the extension below 100 nm in stretching experiments it is possible to unfold the NI6C insert without unfolding the I27 domains. Then, it is possible to relax the partially unfolded construct and measure its refolding force curve. Typically, I27 domains and other beta proteins, once unfolded, relax as simple WLC- like chain and take a few seconds to refold back. However, ankyrin repeats in native proteins [200] as well as in synthetic proteins such as NI6C display an unusual property in that they refold quickly once relaxed and they generate pronounced folding force peaks as shown in Fig. 4C (blue trace). Those force peaks are particularly easy to



ref [23] and (Supp. Info.) capture when the relaxation process is carried out slowly, e.g. at speeds less than 30 nm/s.

8. CG-SMD unfolding and refolding of NI6C (Figs. 4, 5). Fig. 4D shows an example of a CG-SMD simulated force-extension-relaxation trace of NI6C, and the comparisons between AFM measured and computed traces are shown in Fig. 4E and 4F. It is striking that the computed traces show a significant resemblance to the AFM determined force curves. The analysis of unfolding and refolding trajectories (Fig. 5) reveals that the protein unfolding process is vectorial and starts at the C-terminus of NI6C and proceeds to the N-terminus, with force peaks corresponding to the unfolding of individual repeats. The refolding process starts after the chain of unfolded ankyrin repeats relaxes somewhat, and it is initiated by a nucleation event that involves the simultaneous folding of three N-terminal repeats. After the nucleation event, the refolding continues stepwise by adding repeats one by one to the folded nucleus and proceeds in the N to C direction, which is opposite to the unfolding direction. The full movies are uploaded with the paper titled “jbc.M110.179697-2” and “jbc. M110.179697-3”.

6. Exploiting coiled coils probes for identifying unfolding pathways

Since the key information about the directionality of the mechanical unfolding and refolding of NI6C was obtained through computer simulations, it is desirable to verify this prediction using unequivocal experimental approaches. For this purpose, we developed a polypeptide probe that could be inserted at different locations along the protein sequence to produce a characteristic unfolding fingerprint. This fingerprint could be indicative of the unfolding process propagating through these locations in a certain direction and should be consistent with the design of the hybrid protein equipped with the probes.

We chose *anti-parallel* coiled-coil polypeptides [201] as candidates for our mechanical probe as in many proteins, coiled-coil elements seem to be separated structurally from the rest of the protein, suggesting that the insertion of an additional coiled coil element into a host protein, at permissible locations, may introduce a relatively small local structural disturbance. This section of the article describes in detail the choice of coiled coil probes, their construction at the DNA and protein level and provides examples of their use to examine mechanical unfolding pathways of ankyrin repeat proteins.

Fig. 7. Force extension curves of (A) (I27)3-NI10C-(I27)3, (B) (I27)3-NI8CCI2C-(I27)3, (C) (I27)3-NI4CCI6C-(I27)3, (D) (I27)3-NI7CC(SL)I3C-(I27)3 with superimposed individual recordings. On the right are the protein sequences of the different constructs with the respected color gradient, with red representing the N terminus and blue representing the C terminus. Adapted from the Supporting Information to our ref. [23]. (For interpretation of the references to color in this figure

6.1. Coiled coils (CC)

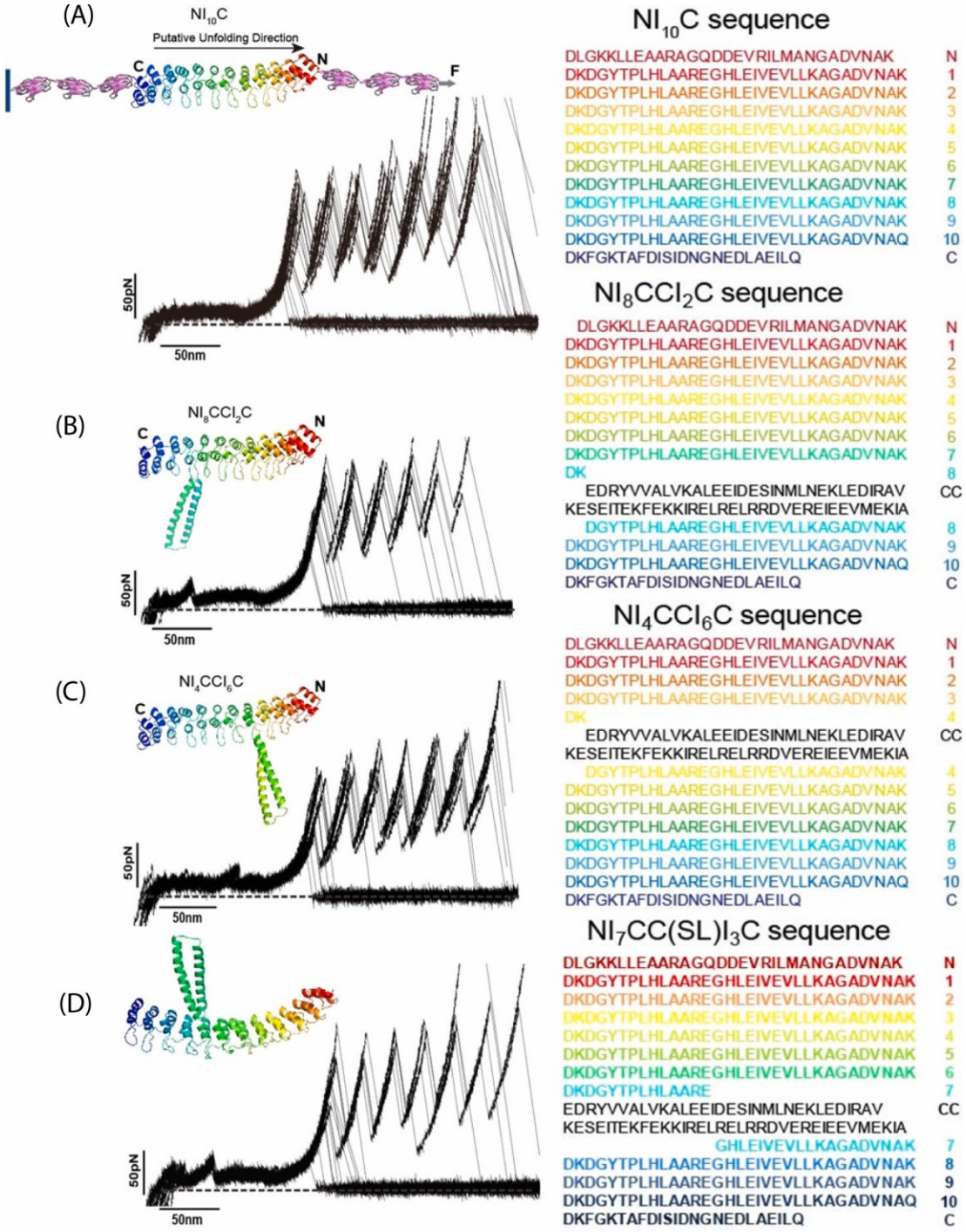
CCs are composed of alpha helical polypeptides forming closely- packed bundles, in which they wrap around each other [201]. Typical CCs are formed by two, three or more helical chains oriented in the same direction (parallel CC) or in the opposite direction (antiparallel CC). Most bundles of parallel and antiparallel CCs are oligomeric as they are composed of several separate polypeptide chains [201]. However, to allow for the facile insertion of a CC probe into the host protein, this needs to be ideally done at the DNA level so the probe becomes an integral part of the protein and does not require any chemical coupling to the host. Thus, the probe needs to encode a single amino acid chain that forms an antiparallel CC in which the C-terminal helix wraps around the N terminal helix. Such antiparallel CC elements exist in natural proteins and an example is shown in Fig. 6A that illustrates the structure of an Archaeal Box C/D sRNP core protein that has a prominent antiparallel CC protruding from the main fold [201]. The fact that this particular CC is pointing away from the main fold and the ends are very close to each other suggested to us that it should be possible to use its sequence to create a similar CC structure in other proteins by inserting the DNA sequence coding for this CC into the DNA sequence of the host protein at chosen locations to create an integral mechanical probe. There are several advantages of using antiparallel, single-chain CCs as compared to other possible probes, e.g. those composed of a structureless polypeptide chain that could also be inserted into the host [148]. For the same number of amino acids, CCs form a more compact structure as compared to a disordered chain and CCs bring the two ends of the insert to a distance as small as 1 nm (as evident in Fig. 6A), minimizing local structural disturbances. In addition, in contrast to unstructured polypeptides that unfold without producing a force peak, CCs such as a leucine zipper were already shown to produce a small but robust unfolding force peak that is easy to capture through SMFS [202,203].

The sequence of our CC probe was determined by investigating the sequence and the 3D structure of the original protein (PDB: 1NT2) and by identifying the two residues at the boundaries of the two coiled alpha helices (residue 82 and residue 151). In our construct, both termini of the CC probe were additionally flanked by Gly-Ser-Gly-Ser linkers to allow some local flexibility so the CC probe, while folding, would not generate excessively strong forces that could locally denature the host protein.

For the study of the directionality of the unfolding of consensus ankyrin repeat proteins, we created a longer variant of the NI6C protein by adding four additional internal repeats to generate a NI10C protein in order to improve our “distance resolution in the protein sequence”.

Since ankyrin repeats are linked by fairly large loops that are oriented perpendicularly to the solenoid structure (Fig. 3, 4A) we chose the sequences coding for those large loops as insertion sites for the CC probe and specifically the regions producing a sharp turn with the distance between neighboring residues close to the distance between the N and C termini of our CC probe. Prolines in ankyrin repeats were excluded as insertion sites since their distinctive cyclic structure of the side chain displays an exceptional conformational rigidity. So, the CC probe was inserted between the Lysine and the Aspartic Acid (residue 45 and 46 in PDB 2QYJ).

In order to verify our hypothesis about the unfolding direction in consensus ankyrin repeat proteins proceeding from the C to the N terminus, we created two hybrid proteins: one with the CC probe integrated close to the C terminus, within the 8th repeat of NI10C, named NI8CCI2C, and one close to the N terminus, within the 4th repeat, named NI4CCI6C. In addition, we created one construct in which the CC



legend, the reader is referred to the web version of this article.)

sequence alone was flanked by three I27 domains in order to perform SMFS measurements to characterize the unfolding fingerprint of our CC probe.

6.2. Mechanical unfolding of the CC probe

We determined the unfolding mechanical fingerprint of the CC from the Archaeal Box C/D sRNP Core Protein by creating a hybrid polyprotein, in which the CC sequence was flanked at the DNA level by very well characterized I27 domains of titin. I27 domains served as SMFS pulling handles and produced their own unique mechanical fingerprint to identify unequivocally single-molecule recordings (Fig. 6). The small, approximately 30 pN peak preceding the unfolding peaks of I27 domains can be attributed to the CC probe. This conjecture is corroborated by the fact that WLC fits before and after this peak produced the contour length increment of 26 nm that corresponds to unfolding of around 70 amino acids, which is close to the length of the probe (67 a.a.).

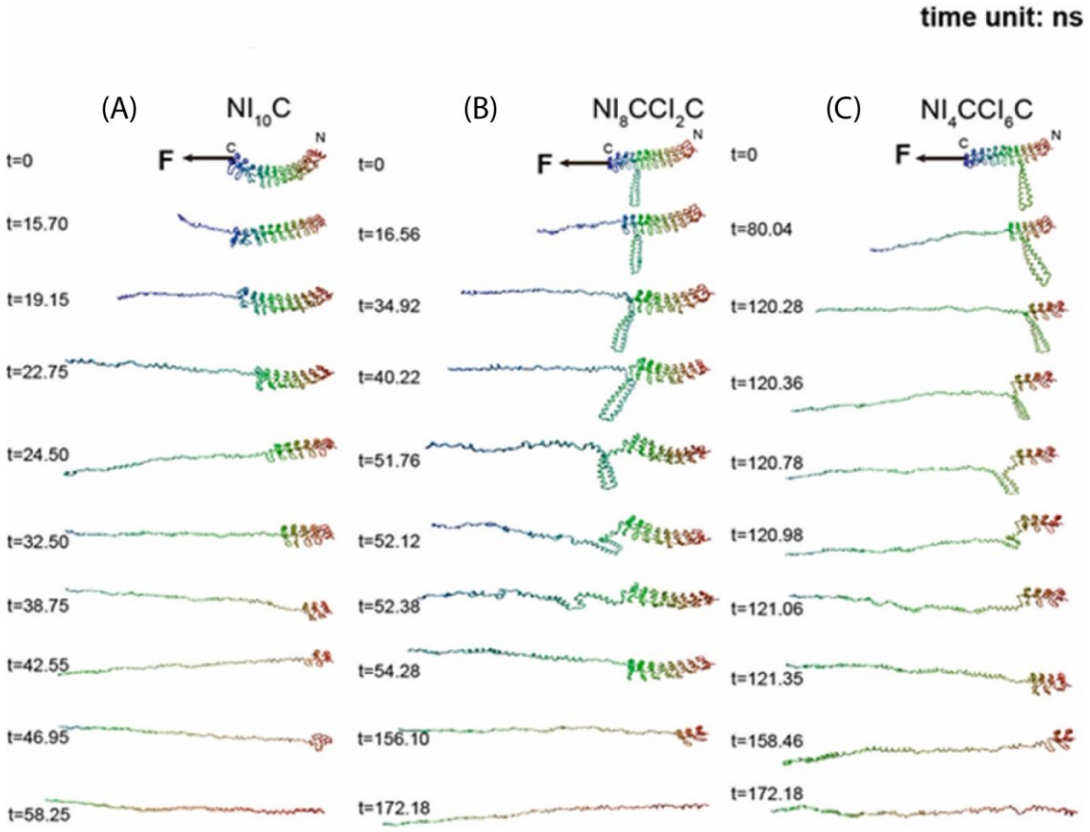


Fig. 8. Snapshots of coarse-grained Steered Molecular Dynamics simulations of NI10C (A), NI8CCI2C (B) and NI4CCI6C (C). NI10C simulation was done with pulling speed of 2.5 $nmns^{-1}$ and spring constant of 6 $pNnm^{-1}$, while NI8CCI2C and NI4CCI6C were performed with pulling speed of 0.5 $nmns^{-1}$ and spring constant of 8 $pNnm^{-1}$. During the simulation, the N termini of all molecules were fixed. Reproduced from the Supporting Information to our ref. [23].

6.3. The force peak of the CC probe captures the progress of the unfolding process

In Fig. 7 we show a set of families of force extension curves of various NI_xCCI_yC constructs, showing that the AFM data is highly reproducible. Even though this superposition smears out individual unfolding peaks of ankyrin repeats, producing an unfolding "plateau" pattern, each graph also distinctly shows an additional peak that may be attributed to the CC probe. Importantly, this peak appears early (at small extensions) in the unfolding pattern of the NI8CCI2C construct (Fig. 7B), and late (at greater extensions) for the NI4CCI6C construct (Fig. 7C), confirming that the unfolding proceeds in the C → N direction. To examine if the specific location within the repeat is important for the probe to capture the unfolding of this region of the protein, we also created an additional construct where the CC probe was inserted in a *small* loop connecting the two helices in the consensus repeat number 8, NI7**CC(SL)**I3C. In this construct, the probe is also located close to the C terminus of the host protein. Force extension recordings for this construct are superimposed in Fig. 7D, and they also reveal the CC peak at an early phase of the unfolding process providing an additional evidence for our hypothesis.

6.4. CG-SMD simulations of the consensus ankyrin repeat proteins equipped with the CC probe

In Fig. 8, we show the summary of our CG-SMD simulations performed on the three consensus ankyrin repeat constructs. They clearly show that similar to NI6C, NI10C also unfolds in the C → N direction. Importantly, the simulations are consistent with our experiments by showing that the proteins equipped with the CC probes also unfold in the C → N direction and that the unfolding of the probe occurs only when the unfolding front progressing through the protein reaches the location of the probe in the construct.

7. Summary

We described the methodologies that combine Atomic Force Microscopy-based SMFS with CG-SMD computer simulations to examine mechanical unfolding and refolding pathways of long repeat proteins. We also described how a compact CC probe can be inserted at various locations in the host protein to capture the progress of the unfolding along its pathway. While here we used ankyrin repeat proteins as a model system, the same procedures can be applied to other repeat proteins (e.g. composed of spectrin, ARM, HEAT and LRR repeats) [204]. More generally, these combined experimental and computational methods worked well with non-repeat multi-domain proteins, such as firefly luciferase [123,124], protein S [205] and the PGK domain [122], suggesting that these approaches should be useful for studying small and large proteins with complex folds and complex unfolding pathways.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work has been supported by a grant from the National Science Foundation (USA), MCB 1817556.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymeth.2021.05.012.

References

[1] T.E. Fisher, P.E. Marszalek, J.M. Fernandez, Stretching single molecules into novel conformations using the atomic force microscope, Nat. Struct. Mol. Biol. 7 (9) (2000) 719–724.

[2] P.E. Marszalek, Y.F. Dufrene, Stretching single polysaccharides and proteins using atomic force microscopy, Chem. Soc. Rev. 41 (9) (2012) 3523–3534.

[3] E.M. Puchner, H.E. Gaub, Force and function: probing proteins with AFM-based force spectroscopy, Curr. Opin. Struct. Biol. 19 (5) (2009) 605–614.

[4] B. Jagannathan, S. Marqusee, Protein folding and unfolding under force, Biopolymers 99 (11) (2013) 860–869.

[5] C. Bustamante, Y.R. Chemla, N.R. Forde, D. Izhaky, Mechanical processes in biochemistry, Annu. Rev. Biochem. 73 (1) (2004) 705–748.

[6] C. Bustamante, J.C. Macosko, G.J.L. Wuite, Grabbing the cat by the tail: Manipulating molecules one by one, Nat. Rev. Mol. Cell Biol. 1 (2) (2000) 130–136.

[7] M.S. Bull, R.M.A. Sullan, H. Li, T.T. Perkins, Improved single molecule force spectroscopy using micromachined cantilevers, ACS Nano 8 (5) (2014) 4984–4995.

[8] M. Hinczewski, C. Hyeon, D. Thirumalai, Directly measuring single-molecule heterogeneity using force spectroscopy, PNAS 113 (27) (2016) E3852–E3861.

[9] C. Hyeon, G. Morrison, D.L. Pincus, D. Thirumalai, Refolding dynamics of stretched biopolymers upon force quench, PNAS 106 (48) (2009) 20288–20293.

[10] P.I. Zhuravlev, M. Hinczewski, S. Chakrabarti, S. Marqusee, D. Thirumalai, Force-dependent switch in protein unfolding pathways and transition-state movements, PNAS 113 (6) (2016) E715–E724.

[11] Y. Zhang, Y. Lyubchenko, The structure of misfolded amyloidogenic dimers: computational analysis of force spectroscopy data, Biophys. J. 107 (12) (2014) 2903–2910.

[12] Z.J. Lv, et al., Mechanism of amyloid beta-protein dimerization determined using single-molecule AFM force spectroscopy, Sci. Rep. 3 (2013).

[13] J.-M. Yuan, C.-L. Chyan, H.-X. Zhou, T.-Y. Chung, H. Peng, G. Ping, G. Yang, The effects of macromolecular crowding on the mechanical stability of protein molecules, Protein Sci. 17 (12) (2008) 2156–2166.

[14] M.C. Williams, I. Rouzina, Force spectroscopy of single DNA and RNA molecules, Curr. Opin. Struct. Biol. 12 (3) (2002) 330–336.

[15] L. Shokri, B. Marintcheva, C.C. Richardson, I. Rouzina, M.C. Williams, Single molecule force spectroscopy of salt-dependent bacteriophage T7 gene 2.5 protein binding to single-stranded DNA, J. Biol. Chem. 281 (50) (2006) 38689–38696.

[16] K.R. Chaurasiya, T. Paramanathan, M.J. McCauley, M.C. Williams, Biophysical characterization of DNA binding from single molecule force measurements, Phys. Life Rev. 7 (3) (2010) 299–341.

[17] A. Galera-Prat, A. Gomez-Sicilia, A.F. Oberhauser, M. Cieplak, M. Carr´ ion-´ Vazquez, Understanding biology by stretching proteins: recent progress, Curr.´ Opin. Struct. Biol. 20 (1) (2010) 63–69.

[18] A.F. Oberhauser, M. Carrion-´ Vazquez, Mechanical biochemistry of proteins one´ molecule at a time, J. Biol. Chem. 283 (11) (2008) 6617–6621.

[19] Y.F. Dufrˆene, E. Evans, A. Engel, J. Helenius, H.E. Gaub, D.J. Müller, Five challenges to bringing single-molecule force spectroscopy into living cells, Nat Meth 8 (2) (2011) 123–127.

[20] P.E. Marszalek, A.F. Oberhauser, Meeting report – NSF-sponsored workshop 'progress and prospects of single-molecule force spectroscopy in biological and chemical sciences', J. Cell Sci. 133 (16) (2020).

[21] Z.N. Scholl, Q. Li, P.E. Marszalek, Single molecule mechanical manipulation for studying biological properties of proteins, DNA, and sugars, WIREs Nanomed. Nanobiotechnol. 6 (3) (2014) 211–229.

[22] W. Lee, X. Zeng, H.-X. Zhou, V. Bennett, W. Yang, P.E. Marszalek, Full reconstruction of a vectorial protein folding pathway by atomic force microscopy and molecular dynamics simulations, J. Biol. Chem. 285 (49) (2010) 38167–38172.

[23] Q. Li, Z.N. Scholl, P.E. Marszalek, Capturing the mechanical unfolding pathway of a large protein with coiled-coil probes, Angew. Chem.-Int. Ed. 53 (49) (2014) 13429–13433.

[24] G. Binnig, C.F. Quate, C.h. Gerber, Atomic force microscope, Phys. Rev. Lett. 56 (9) (1986) 930–933.

[25] S. Posch, C. Aponte-Santamaría, R. Schwarzl, A. Karner, M. Radtke, F. Grater,¨ T. Obser, G. Konig, M.A. Brehm, H.J. Gruber, R.R. Netz, C. Baldauf,¨ R. Schneppenheim, R. Tamp´e, P. Hinterdorfer, Mutual A domain interactions in the force sensing protein von Willebrand factor, J. Struct. Biol. 197 (1) (2017) 57–64.

[26] A. Xiao, H. Li, Direct monitoring of equilibrium protein folding-unfolding by atomic force microscopy: pushing the limit, Chem. Commun. 55 (86) (2019) 12920–12923.

[27] F. Kienberger, A. Ebner, H.J. Gruber, P. Hinterdorfer, Molecular recognition imaging and force spectroscopy of single biomolecules, Acc. Chem. Res. 39 (1) (2006) 29–36.

[28] H. Yu, M.G.W. Siewny, D.T. Edwards, A.W. Sanders, T.T. Perkins, Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins, Science 355 (6328) (2017) 945–950.

[29] A. Ashkin, J. Dziedzic, Optical trapping and manipulation of viruses and bacteria, Science 235 (4795) (1987) 1517–1520.

[30] A. Ashkin, J.M. Dziedzic, T. Yamane, Optical trapping and manipulation of single cells using infrared laser beams, Nature 330 (1987) 769.

[31] C. Bustamante, et al., Single-Molecule Studies of Protein Folding with Optical Tweezers, in Annual Review of Biochemistry, Vol 89, R.D. Kornberg, Editor. 2020. p. 443-470.

[32] X. Chen, N. Rajasekaran, K. Liu, C.M. Kaiser, Synthesis runs counter to directional folding of a nascent protein domain, Nat. Commun. 11 (1) (2020), https://doi.org/10.1038/s41467-020-18921-8.

[33] K. Liu, X. Chen, C.M. Kaiser, Energetic dependencies dictate folding mechanism in a complex protein, PNAS 116 (51) (2019) 25641–25648.

[34] H. Motlagh, D. Toptygin, C. Kaiser, V. Hilser, Single-molecule chemo-mechanical spectroscopy provides structural identity of folding intermediates, Biophys. J. 110 (6) (2016) 1280–1290.

[35] Z. Ganim, M. Rief, Mechanically switching single-molecule fluorescence of GFP by unfolding and refolding, PNAS 114 (42) (2017) 11052–11056.

[36] S.S. Mandal, D.R. Merz, M. Buchsteiner, R.I. Dima, M. Rief, G. Zol ˇ dak,´ Nanomechanics of the substrate binding domain of Hsp70 determine its allosteric ATP-induced conformational change, PNAS 114 (23) (2017) 6040–6045.

[37] S. Meinhold, D. Bauer, J. Huber, U. Merkel, A. Weißl, G. Zol ˇ dak, M. Rief, An´ active, ligand-responsive pulling geometry reports on internal signaling between subdomains of the DnaK nucleotide-binding domain in single-molecule mechanical experiments, Biochemistry 58 (47) (2019) 4744–4750.

[38] J.R. Moffitt, et al., Recent advances in optical tweezers, in Annual Review of Biochemistry. 2008, Annual Reviews: Palo Alto. p. 205–228.

[39] B.A. Berghuis, M. Kober, T. van Laar, N.H. Dekker, High-throughput, high-force¨ probing of DNA-protein interactions with magnetic tweezers, Methods 105 (2016) 90–98.

[40] K.C. Neuman, A. Nagy, Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy, Nat. Methods 5 (2008) 491.

[41] I. Popa, J.A. Rivas-Pardo, E.C. Eckels, D.J. Echelman, C.L. Badilla, J. Valle-Orero, J.M. Fernandez, A HaloTag anchored ruler for week-long studies of protein´ dynamics, J. Am. Chem. Soc. 138 (33) (2016) 10546–10553.

[42] S. Smith, L. Finzi, C. Bustamante, Direct mechanical measurements of the elasticity of single DNA-molecules by using magnetic beads, Science 258 (5085) (1992) 1122–1126.

[43] F. Kriegel, N. Ermann, J. Lipfert, Probing the mechanical properties, conformational changes, and interactions of nucleic acids with magnetic tweezers, J. Struct. Biol. 197 (1) (2017) 26–36.

[44] I. De Vlaminck, C. Dekker, Recent advances in magnetic tweezers, in: D.C. Rees (Ed.), Annual Review of Biophysics, Vol. 41, Annual Reviews, Palo Alto, 2012, pp. 453–472.

[45] J.P. Renn, S. Bhattacharyya, H. Bai, C. He, H. Li, A.F. Oberhauser, J.F. Marko, D. E. Makarov, A. Matouschek, Mechanical unfolding of spectrin reveals a super-exponential dependence of unfolding rate on force, Sci. Rep. 9 (1) (2019), https://doi.org/10.1038/s41598-019-46525-w.

[46] A.F. Oberhauser, P.K. Hansma, M. Carrion-Vazquez, J.M. Fernandez, Stepwise unfolding of titin under force-clamp atomic force microscopy, PNAS 98 (2) (2001) 468–472.

[47] H. Clausen-Schaumann, M. Rief, H.E. Gaub, Sequence dependent mechanics of single DNA molecules. Biophys. J., 1999. 76(1): p. A151–A151.

[48] H. Clausen-Schaumann, M. Rief, C. Tolksdorf, H.E. Gaub, Mechanical stability of single DNA molecules, Biophys. J. 78 (4) (2000) 1997–2007.

[49] C. Bustamante, Z. Bryant, S.B. Smith, Ten years of tension: single-molecule DNA mechanics, Nature 421 (6921) (2003) 423–427.

[50] J. Gore, Z. Bryant, M. Nollmann, M.U. Le, N.R. Cozzarelli, C. Bustamante, DNA¨ overwinds when stretched, Nature 442 (7104) (2006) 836–839.

[51] S.B. Smith, Y. Cui, C. Bustamante, Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules, Science 271 (5250) (1996) 795–799.

[52] G.J.L. Wuite, S.B. Smith, M. Young, D. Keller, C. Bustamante, Single-molecule studies of the effect of template tension on T7 DNA polymerase activity, Nature 404 (6773) (2000) 103–106.

[53] Z.N. Scholl, M. Rabbi, D. Lee, L. Manson, H. S-Gracz, P.E. Marszalek, Origin of overstretching transitions in single-stranded nucleic acids, Phys. Rev. Lett. 111 (18) (2013), https://doi.org/10.1103/PhysRevLett.111.188302.

[54] C. Ke, M. Humeniuk, H. S-Gracz, P.E. Marszalek, Direct Measurements of base stacking interactions in DNA by single-molecule atomic-force spectroscopy, Phys. Rev. Lett. 99 (1) (2007), https://doi.org/10.1103/PhysRevLett.99.018302.

[55] C. Bustamante, J. Marko, E. Siggia, S. Smith, Entropic elasticity of [lambda]- phage DNA, Science 265 (5178) (1994) 1599–1600.

[56] Z. Bryant, M.D. Stone, J. Gore, S.B. Smith, N.R. Cozzarelli, C. Bustamante, Structural transitions and elasticity from torque measurements on DNA, Nature 424 (6946) (2003) 338–341.

[57] C. Bustamante, W. Cheng, Y.X. Mejia, Revisiting the central dogma one molecule at a time, Cell 144 (4) (2011) 480–497.

[58] R.J. Davenport, et al., Single-molecule study of transcriptional pausing and arrest by E-coli RNA polymerase, Science 287 (5462) (2000) 2497–2500.

[59] J. Gore, Z. Bryant, M.D. Stone, M. Nollmann, N.R. Cozzarelli, C. Bustamante,¨ Mechanochemical analysis of DNA gyrase using rotor bead tracking, Nature 439 (7072) (2006) 100–104.

[60] G. Charvin, et al., Tracking topoisomerase activity at the single-molecule level, in Annual Review of Biophysics and Biomolecular Structure. 2005, Annual Reviews: Palo Alto. p. 201–219.

[61] T.R. Strick, V. Croquette, D. Bensimon, Single-molecule analysis of DNA uncoiling by a type II topoisomerase, Nature 404 (6780) (2000) 901–904.

[62] G. Witz, G. Dietler, A. Stasiak, Tightening of DNA knots by supercoiling facilitates their unknotting by type II DNA topoisomerases, Proc. Natl. Acad. Sci. 108 (9) (2011) 3608–3611.

[63] H.B. Li, et al., Single-molecule force spectroscopy on polysaccharides by AFM – nanomechanical fingerprint of alpha-(1,4)-linked polysaccharides, Chem. Phys. Lett. 305 (3–4) (1999) 197–201.

[64] M. Rief, et al., Single molecule force spectroscopy on polysaccharides by atomic force microscopy, Science 275 (5304) (1997) 1295–1297.

[65] G. Lee, W. Nowak, J. Jaroniec, Q. Zhang, P.E. Marszalek, Nanomechanical control of glucopyranose rotamers, J. Am. Chem. Soc. 126 (20) (2004) 6218–6219.

[66] W. Lee, et al., Mechanics of Polysaccharides, in Molecular Manipulation with Atomic Force Microscopy. N.W. Anne-Sophie Duwez, Editor. 2012, CRC Press: Boca Raton London New York.

[67] P.E. Marszalek, A.F. Oberhauser, Y.-P. Pang, J.M. Fernandez, Polysaccharide elasticity governed by chair-boat transitions of the glucopyranose ring, Nature 396 (6712) (1998) 661–664.

[68] P.E. Marszalek, Y.-P. Pang, H. Li, J.E. Yazal, A.F. Oberhauser, J.M. Fernandez, Atomic levers control pyranose ring conformations, PNAS 96 (14) (1999) 7894–7898.

[69] Q. Zhang, P.E. Marszalek, Identification of sugar isomers by single-molecule force spectroscopy, J. Am. Chem. Soc. 128 (17) (2006) 5596–5597.

[70] H. Li, W.A. Linke, A.F. Oberhauser, M. Carrion-Vazquez, J.G. Kerkvliet, H. Lu, P. E. Marszalek, J.M. Fernandez, Reverse engineering of the giant muscle protein titin, Nature 418 (6901) (2002) 998–1002.

[71] Y. Cao, H. Li, Single molecule force spectroscopy reveals a weakly populated microstate of the FnIII domains of tenascin, J. Mol. Biol. 361 (2) (2006) 372–381.

[72] Y. Cao, H. Li, Polyprotein of GB1 is an ideal artificial elastomeric protein, Nat. Mater. 6 (2) (2007) 109–114.

[73] H. Li, Y. Cao, Protein mechanics: from single molecules to functional biomaterials, Acc. Chem. Res. 43 (10) (2010) 1331–1341.

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

[75] M. Schlierf, H. Li, J.M. Fernandez, The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques, PNAS 101 (19) (2004) 7299–7304.

[76] D. Sharma, O. Perisic, Q. Peng, Y. Cao, C. Lam, H. Lu, H. Li, Single-molecule force spectroscopy reveals a mechanically stable protein fold and the rational tuning of its mechanical stability, Proc. Natl. Acad. Sci. 104 (22) (2007) 9278–9283.

[77] A. Valbuena, A. Vera, J. Oroz, M. Men´endez, M. Carrion-´ Vazquez, Mechanical´ properties of β-catenin revealed by single-molecule experiments, Biophys. J . 103 (8) (2012) 1744–1752.

[78] T. Isgro, et al., NAMD Tutorial, in Unix/MacOSX Version 2009: Urbana.

[79] W.A. Linke, A. Grützner, Pulling single molecules of titin by AFM—recent advances and physiological implications, Pflügers Archiv Eur. J. Physiol. 456 (1) (2008) 101–115.

[80] V. Moy, E. Florin, H. Gaub, Intermolecular Forces and Energies between Ligands and Receptors, Science 266 (5183) (1994) 257–259.

[81] V.T. Moy, E.-L. Florin, H.E. Gaub, Adhesive forces between ligand and receptor measured by afm, Colloids and Surf. A 93 (1994) 343–348.

[82] E. Florin, V. Moy, H. Gaub, Adhesion forces between individual ligand-receptor pairs, Science 264 (5157) (1994) 415–417.

[83] F. Rico, C. Chu, M.H. Abdulreda, Y. Qin, V.T. Moy, Temperature modulation of integrin-mediated cell adhesion, Biophys. J. 99 (5) (2010) 1387–1396.

[84] Y.F. Dufrˆene, P. Hinterdorfer, Recent progress in AFM molecular recognition studies, Pflügers Archiv Eur. J. Physiol. 456 (1) (2008) 237–245.

[85] S. Katletz, C. Stroh, C. Rankl, U.M. Titulaer, P. Hinterdorfer, Force-induced lysozyme–HyHEL5 antibody dissociation and its analysis by means of a cooperative binding model, Biophys. J. 99 (1) (2010) 323–332.

[86] J. Tang, A. Ebner, N. Ilk, H. Lichtblau, C. Huber, R. Zhu, D. Pum, M. Leitner, V. Pastushenko, H.J. Gruber, U.B. Sleytr, P. Hinterdorfer, High-affinity tags fused to S-layer proteins probed by atomic force microscopy, Langmuir 24 (4) (2008) 1324–1329.

[87] J. Tang, D. Krajcikova, R. Zhu, A. Ebner, S. Cutting, H.J. Gruber, I. Barak, P. Hinterdorfer, Atomic force microscopy imaging and single molecule recognition force spectroscopy of coat proteins on the surface of Bacillus subtilis spore, J. Mol. Recognit. 20 (6) (2007) 483–489.

[88] D.J. Müller, J. Helenius, D. Alsteens, Y.F. Dufrˆene, Force probing surfaces of living cells to molecular resolution, Nat. Chem. Biol. 5 (6) (2009) 383–390.

[89] M. Krieg, G. Flaschner, D. Alsteens, B.M. Gaub, W.H. Roos, G.J.L. Wuite, H.¨ E. Gaub, C. Gerber, Y.F. Dufrˆene, D.J. Müller, Atomic force microscopy-based mechanobiology, Nat. Rev. Phys. 1 (1) (2019) 41–57.

[90] L.F. Milles, E.M. Unterauer, T. Nicolaus, H.E. Gaub, Calcium stabilizes the strongest protein fold, Nat. Commun. 9 (1) (2018), https://doi.org/10.1038/ s41467-018-07145-6.

[91] A. Ebner, F. Kienberger, C. Huber, A.S.M. Kamruzzahan, V.P. Pastushenko, J. Tang, G. Kada, H.J. Gruber, U.B. Sleytr, M. Sara, P. Hinterdorfer, Atomic-force-´ microscopy imaging and molecular-recognition-force microscopy of recrystallized heterotetramers comprising an S-layer-streptavidin fusion protein, ChemBioChem 7 (4) (2006) 588–591.

[92] P. Hinterdorfer, Y.F. Dufrene, Detection and localization of single molecularˆ recognition events using atomic force microscopy, Nat Meth 3 (5) (2006) 347–355.

[93] N. Willet, et al., Molecular Recognition Force Spectroscopy, in Molecular Manipulation with Atomic Force Microscopy. N.W. Anne-Sophie Duwez, Editor. 2012, CRC Press: Boca Raton London New York.

[94] M. Jahn, K. Tych, H. Girstmair, M. Steinmaßl, T. Hugel, J. Buchner, M. Rief, Folding and domain interactions of three orthologs of Hsp90 studied by single- molecule force spectroscopy, Structure 26 (1) (2018) 96–105.e4.

[95] M. Mickler, R.I. Dima, H. Dietz, C. Hyeon, D. Thirumalai, M. Rief, Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations, PNAS 104 (51) (2007) 20268–20273.

[96] X. Zhuang, M. Rief, Single-molecule folding, Curr. Opin. Struct. Biol. 13 (1) (2003) 88–97.

[97] G. Zoldak, J. Stigler, B. Pelz, H. Li, M. Rief, Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy, Proc. Natl. Acad. Sci. 110 (45) (2013) 18156–18161.

[98] D.T. Edwards, et al., Force spectroscopy with 9-μs resolution and sub-pN stability achieved by tailoring AFM cantilever geometry, Biophys. J. (2017).

[99] C.Z. He, et al., Direct observation of the reversible two-state unfolding and refolding of an alpha/beta protein by single-molecule atomic force microscopy, Angew. Chem.-Int. Ed. 54 (34) (2015) 9921–9925.

[100] R. Walder, et al., Rapid characterization of a mechanically labile α-helical protein enabled by efficient site-specific bioconjugation, J. Am. Chem. Soc. (2017).

[101] W. Hu, B.T. Walters, Z.-Y. Kan, L. Mayne, L.E. Rosen, S. Marqusee, S. W. Englander, Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry, Proc. Natl. Acad. Sci. 110 (19) (2013) 7684–7689.

[102] M.K. Jensen, A.J. Samelson, A. Steward, J. Clarke, S. Marqusee, The folding and unfolding behavior of ribonuclease H on the ribosome, J. Biol. Chem. 295 (33) (2020) 11410–11417.

[103] E.A. Shank, C. Cecconi, J.W. Dill, S. Marqusee, C. Bustamante, The folding cooperativity of a protein is controlled by its chain topology, Nature 465 (7298) (2010) 637–640.

[104] S. Batey, K.A. Scott, J. Clarke, Complex folding kinetics of a multidomain protein, Biophys. J. 90 (6) (2006) 2120–2130.

[105] R.B. Best, D.J. Brockwell, J.L. Toca-Herrera, A.W. Blake, D.A. Smith, S.E. Radford, J. Clarke, Force mode atomic force microscopy as a tool for protein folding studies, Anal. Chim. Acta 479 (1) (2003) 87–105.

[106] A. Borgia, P.M. Williams, J. Clarke, Single-molecule studies of protein folding, in Annual Review of Biochemistry. 2008, Annual Reviews: Palo Alto. p. 101–125.

[107] J.-H. Han, S. Batey, A.A. Nickson, S.A. Teichmann, J. Clarke, The folding and evolution of multidomain proteins, Nat. Rev. Mol. Cell Biol. 8 (4) (2007) 319–330.

[108] S. Ng, et al., Single molecule studies of protein folding by atomic force microscopy (AFM). Abstracts of Papers of the American Chemical Society, 2004. 227: p. U545–U545.

[109] O.B. Nilsson, A.A. Nickson, J.J. Hollins, S. Wickles, A. Steward, R. Beckmann, G. von Heijne, J. Clarke, Cotranslational folding of spectrin domains via partially structured states, Nat. Struct. Mol. Biol. 24 (3) (2017) 221–225.

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

[111] K.A. Scott, et al., Titin; a multidomain protein that behaves as the sum of its parts, J. Mol. Biol. 315 (4) (2002) 819–829.

[112] P.M. Williams, S.B. Fowler, R.B. Best, J. Luis Toca-Herrera, K.A. Scott, A. Steward, J. Clarke, Hidden complexity in the mechanical properties of titin, Nature 422 (6930) (2003) 446–449.

[113] D.H. Goldman, et al., Mechanical force releases nascent chain–mediated ribosome arrest in vitro and in vivo, Science (New York, N.Y.) 348 (6233) (2015) 457–460.

[114] C.M. Kaiser, D.H. Goldman, J.D. Chodera, I. Tinoco, C. Bustamante, The ribosome modulates nascent protein folding, Science 334 (6063) (2011) 1723–1727.

[115] M. Carrion-Vazquez, A.F. Oberhauser, S.B. Fowler, P.E. Marszalek, S.E. Broedel, J. Clarke, J.M. Fernandez, Mechanical and chemical unfolding of a single protein: a comparison, Proc. Natl. Acad. Sci. U.S.A. 96 (7) (1999) 3694–3699.

[116] J.M. Fernandez, H. Li, Force-clamp spectroscopy monitors the folding trajectory of a single protein, Science 303 (5664) (2004) 1674–1678.

[117] S. Garcia-Manyes, et al., Single-molecule force spectroscopy predicts a misfolded, domain-swapped conformation in human D-crystallin protein, J. Biol. Chem. 291 (8) (2016) 4226–4235.

[118] A.F. Oberhauser, P.E. Marszalek, H.P. Erickson, J.M. Fernandez, The molecular elasticity of the extracellular matrix protein tenascin, Nature 393 (6681) (1998) 181–185.

[119] I. Popa, R. Berkovich, J. Alegre-Cebollada, J.A. Rivas-Pardo, J.M. Fernandez, Halotag tethers to study titin folding at the single molecule level, Biophys. J . 106 (2) (2014) 391a, https://doi.org/10.1016/j.bpj.2013.11.2211.

[120] J. Valle-Orero, J.A. Rivas-Pardo, R. Tapia-Rojo, I. Popa, D.J. Echelman, S. Haldar, J.M. Fernandez, Mechanical deformation accelerates protein ageing, Angew.´ Chem.-Int. Ed. 56 (33) (2017) 9741–9746.

[121] J. Valle-Orero, R. Tapia-Rojo, E.C. Eckels, J.A. Rivas-Pardo, I. Popa, J. M. Fernandez, Proteins breaking bad: a free energy perspective, J. Phys. Chem.´ Lett. 8 (15) (2017) 3642–3647.

[122] Q. Li, Z.N. Scholl, P.E. Marszalek, Unraveling the mechanical unfolding pathways of a multidomain protein: phosphoglycerate kinase, Biophys. J. 115 (1) (2018) 46–58.
[123] Z.N. Scholl, W. Yang, P.E. Marszalek, Chaperones rescue luciferase folding by separating its domains, J. Biol. Chem. 289 (41) (2014) 28607–28618.
[124] Z.N. Scholl, W. Yang, P.E. Marszalek, Competing pathways and multiple folding nuclei in a large multidomain protein, Luciferase, Biophys. J. 112 (9) (2017) 1829–1840.
[125] T.R. Strick, M.-N. Dessinges, G. Charvin, N.H. Dekker, J.-F. Allemand, D. Bensimon, V. Croquette, Stretching of macromolecules and proteins, Rep. Prog. Phys. 66 (1) (2003) 1–45.
[126] Y. Cao, R. Kuske, H. Li, Direct observation of markovian behavior of the mechanical unfolding of individual proteins, Biophys. J. 95 (2) (2008) 782–788. [127] H. Li, A.F. Oberhauser, S.D. Redick, M. Carrion-Vazquez, H.P. Erickson, J. M. Fernandez, Multiple conformations of PEVK proteins detected by single- molecule techniques, PNAS 98 (19) (2001) 10682–10686.
[128] P.E. Marszalek, H. Lu, H. Li, M. Carrion-Vazquez, A.F. Oberhauser, K. Schulten, J. M. Fernandez, Mechanical unfolding intermediates in titin modules, Nature 402 (6757) (1999) 100–103.
[129] B. Jagannathan, P.J. Elms, C. Bustamante, S. Marqusee, Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein, Proc. Natl. Acad. Sci. 109 (44) (2012) 17820–17825.
[130] C.P. Johnson, H.-Y. Tang, C. Carag, D.W. Speicher, D.E. Discher, Forced unfolding of proteins within cells, Science 317 (5838) (2007) 663–666.
[131] R. Law, et al., Cooperativity in forced unfolding of tandem spectrin repeats, Biophys. J. 84 (1) (2003) 533–544.
[132] H. Lei, C. He, C. Hu, J. Li, X. Hu, X. Hu, H. Li, Single-molecule force spectroscopy trajectories of a single protein and its polyproteins are equivalent: a direct experimental validation based on a small protein NuG2, Angew. Chem.-Int. Ed. 56 (22) (2017) 6117–6121.
[133] L. Li, S. Wetzel, A. Plückthun, J.M. Fernandez, Stepwise unfolding of ankyrin repeats in a single protein revealed by atomic force microscopy, Biophys. J. 90 (4) (2006) L30–L32.
[134] R. Zhu, W. Sandtner, J.E.A. Ahiable, A.H. Newman, M. Freissmuth, H.H. Sitte, P. Hinterdorfer, Allosterically linked binding sites in serotonin transporter revealed by single molecule force spectroscopy, Front. Mol. Biosci. 7 (2020), https://doi.org/10.3389/fmolb.2020.00099.
[135] M. Schlierf, F. Berkemeier, M. Rief, Direct observation of active protein folding using lock-in force spectroscopy, Biophys. J. 93 (11) (2007) 3989–3998.
[136] I. Schwaiger, A. Kardinal, M. Schleicher, A.A. Noegel, M. Rief, A mechanical unfolding intermediate in an actin-crosslinking protein, Nat. Struct. Mol. Biol. 11 (1) (2004) 81–85.
[137] I. Schwaiger, M. Schleicher, A.A. Noegel, M. Rief, The folding pathway of a fast- folding immunoglobulin domain revealed by single-molecule mechanical experiments, EMBO Rep. 6 (1) (2005) 46–51.
[138] J. Stigler, F. Ziegler, A. Gieseke, J.C.M. Gebhardt, M. Rief, The complex folding network of single calmodulin molecules, Science 334 (6055) (2011) 512–516.
[139] M.S.Z. Kellermayer, et al., Folding-unfolding transitions in single titin molecules characterized with laser tweezers, Science 276 (5315) (1997) 1112–1116.
[140] L. Tskhovrebova, J. Trinick, J.A. Sleep, R.M. Simmons, Elasticity and unfolding of single molecules of the giant muscle protein titin, Nature 387 (6630) (1997) 308–312.
[141] M. Rief, et al., Reversible unfolding of individual titin immunoglobulin domains by AFM, Science 276 (5315) (1997) 1109–1112.
[142] H. Dietz, F. Berkemeier, M. Bertz, M. Rief, Anisotropic deformation response of single protein molecules, PNAS 103 (34) (2006) 12724–12728.
[143] H. Dietz, M. Bertz, M. Schlierf, F. Berkemeier, T. Bornschlogl, J.P. Junker, M. Rief,¨ Cysteine engineering of polyproteins for single-molecule force spectroscopy, Nat. Protocols 1 (1) (2006) 80–84.
[144] H. Dietz, M. Rief, Protein structure by mechanical triangulation, PNAS 103 (5) (2006) 1244–1247.
[145] E. Pfitzner, C. Wachauf, F. Kilchherr, B. Pelz, W.M. Shih, M. Rief, H. Dietz, Rigid DNA beams for high-resolution single-molecule mechanics, Angew. Chem. Int. Ed. 52 (30) (2013) 7766–7771.
[146] H. Dietz, M. Rief, Exploring the energy landscape of GFP by single-molecule mechanical experiments, PNAS 101 (46) (2004) 16192–16197.
[147] M. Hinczewski, J.C.M. Gebhardt, M. Rief, D. Thirumalai, From mechanical folding trajectories to intrinsic energy landscapes of biopolymers, Proc. Natl. Acad. Sci. 110 (12) (2013) 4500–4505.
[148] M. Carrion-Vazquez, P.E. Marszalek, A.F. Oberhauser, J.M. Fernandez, Atomic force microscopy captures length phenotypes in single proteins, PNAS 96 (20) (1999) 11288–11292.
[149] M. Rabbi, P. Marszalek, Probing polysaccharide and protein mechanics by atomic force microscopy, in: P.R. Selvin, T. Ha (Eds.), Single-Molecule Techniques: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2008, pp. 371–394.
[150] Z.N. Scholl, P.E. Marszalek, Improving single molecule force spectroscopy through automated real-time data collection and quantification of experimental conditions, Ultramicroscopy 136 (2014) 7–14.
[151] Z.N. Scholl, Q. Li, E. Josephs, D. Apostolidou, P.E. Marszalek, Force spectroscopy of single protein molecules using an atomic force microscope, JoVE (144) (2019), https://doi.org/10.3791/55989.
[152] D.T. Edwards, T.T. Perkins, Optimizing force spectroscopy by modifying commercial cantilevers: Improved stability, precision, and temporal resolution, J. Struct. Biol. 197 (1) (2017) 13–25.
[153] J.K. Faulk, et al., Improved force spectroscopy using focused-ion-beam-modified cantilevers, in: M. Spies, Y.R. Chemla (Eds.), Single-Molecule Enzymology: Nanomechanical Manipulation and Hybrid Methods, Elsevier Academic Press Inc, San Diego, 2017, pp. 321–351.
[154] R.M.A. Sullan, A.B. Churnside, D.M. Nguyen, M.S. Bull, T.T. Perkins, Atomic force microscopy with sub-picoNewton force stability for biological applications, Methods 60 (2) (2013) 131–141.
[155] K. Pawlak, J. Strzelecki, Nanopuller-open data acquisition platform for AFM force spectroscopy experiments, Ultramicroscopy 164 (2016) 17–23.
[156] E.-L. Florin, M. Rief, H. Lehmann, M. Ludwig, C. Dornmair, V.T. Moy, H.E. Gaub, Sensing specific molecular-interactions with the atomic-force microscope, Biosens. Bioelectron. 10 (9-10) (1995) 895–901.
[157] H.-J. Butt, B. Cappella, M. Kappl, Force measurements with the atomic force microscope: technique, interpretation and applications, Surf. Sci. Rep. 59 (1-6) (2005) 1–152.
[158] H.-J. Butt, M. Jaschke, Calculation of thermal noise in atomic force microscopy, Nanotechnology 6 (1) (1995) 1–7.
[159] A.B. Churnside, R.M.A. Sullan, D.M. Nguyen, S.O. Case, M.S. Bull, G.M. King, T. T. Perkins, Routine and timely sub-piconewton force stability and precision for biological applications of atomic force microscopy, Nano Lett. 12 (7) (2012) 3557–3561.
[160] M.-A. LeBlanc, D.T. Edwards, R. Walder, D. Rabuka, T.T. Perkins, M.C. Sousa, Heterobifunctional polyprotein for efficient characterization of mechanically labile proteins, Biophys. J. 112 (3) (2017) 455a–456a.
[161] J. Li, H. Li, Single molecule force spectroscopy reveals that a two-coordinate ferric site is critical for the folding of holo-rubredoxin, Nanoscale 12 (44) (2020) 22564–22573.
[162] A. Ebner. Functionalization of probe tips and supports for single-molecule recognition force Microscopy, in Stm and Afm Studies On, P. Samori, Editor. 2008, Springer-Verlag Berlin: Berlin. p. 29–76.
[163] T.D. Becke, S. Ness, S. Sudhop, H.E. Gaub, M. Hilleringmann, A.F. Schilling, H. Clausen-Schaumann, Covalent immobilization of proteins for the single molecule force spectroscopy, Jove-J. Visual. Exp. (138) (2018), https://doi.org/ 10.3791/58167.
[164] I. Popa, R. Berkovich, J. Alegre-Cebollada, C.L. Badilla, J.A. Rivas-Pardo, Y. Taniguchi, M. Kawakami, J.M. Fernandez, Nanomechanics of HaloTag tethers, J. Am. Chem. Soc. 135 (34) (2013) 12762–12771.
[165] I. Popa, et al., Nano-mechanics of HaloTag tethers, J. Am. Chem. Soc. (2013).
[166] M. Otten, W. Ott, M.A. Jobst, L.F. Milles, T. Verdorfer, D.A. Pippig, M.A. Nash, H. E. Gaub, From genes to protein mechanics on a chip, Nat. Methods 11 (11) (2014) 1127–1130.
[167] S.W. Stahl, M.A. Nash, D.B. Fried, M. Slutzki, Y. Barak, E.A. Bayer, H.E. Gaub, Single-molecule dissection of the high-affinity cohesin–dockerin complex, Proc. Natl. Acad. Sci. 109 (50) (2012) 20431–20436.
[168] C. Schoeler, K.H. Malinowska, R.C. Bernardi, L.F. Milles, M.A. Jobst, E. Durner, W. Ott, D.B. Fried, E.A. Bayer, K. Schulten, H.E. Gaub, M.A. Nash, Ultrastable cellulosome-adhesion complex tightens under load, Nat. Commun. 5 (1) (2014), https://doi.org/10.1038/ncomms6635.
[169] C. Schoeler, R.C. Bernardi, K.H. Malinowska, E. Durner, W. Ott, E.A. Bayer, K. Schulten, M.A. Nash, H.E. Gaub, Mapping mechanical force propagation through biomolecular complexes, Nano Lett. 15 (11) (2015) 7370–7376.
[170] R. Yang, Y.H. Wong, G.K.T. Nguyen, J.P. Tam, J. Lescar, B. Wu, Engineering a catalytically efficient recombinant protein ligase, J. Am. Chem. Soc. 139 (15) (2017) 5351–5358.
[171] W. Ott, E. Durner, H.E. Gaub, Enzyme-mediated, site-specific protein coupling strategies for surface-based binding assays, Angew. Chem.-Int. Ed. 57 (39) (2018) 12666–12669.
[172] Y. Deng, T. Wu, M. Wang, S. Shi, G. Yuan, X.i. Li, H. Chong, B. Wu, P. Zheng, Enzymatic biosynthesis and immobilization of polyprotein verified at the single- molecule level, Nat. Commun. 10 (1) (2019), https://doi.org/10.1038/s41467- 019-10696-x.
[173] R. Satija, A.M. Berezhkovskii, D.E. Makarov, Broad distributions of transition- path times are fingerprints of multidimensionality of the underlying free energy landscapes, PNAS 117 (44) (2020) 27116–27123.
[174] T. Merz, S.K. Wetzel, S. Firbank, A. Plückthun, M.G. Grütter, P.R.E. Mittl, Stabilizing ionic interactions in a full-consensus ankyrin repeat protein, J. Mol. Biol. 376 (1) (2008) 232–240.
[175] A. Steward, J.L. Toca-Herrera, J. Clarke, Versatile cloning system for construction of multimeric proteins for use in atomic force microscopy, Protein Sci. 11 (9) (2002) 2179–2183.
[176] M. Gao et al., Identifying unfolding intermediates of FN-III-10 with steered molecular dynamics simulations. Biophys. J., 2002. 82(1): p. 133a–133a.
[177] M. Gao et al., Steered molecular dynamics simulation of the unfolding of multiple titin immunoglobulin domains. Biophys. J., 2000. 78(1): p. 28a–28a.
[178] B. Isralewitz, J. Baudry, J. Gullingsrud, D. Kosztin, K. Schulten, Steered molecular dynamics investigations of protein function, J. Mol. Graph. Model. 19 (1) (2001) 13–25.
[179] W. Lee, et al., Mutation of conserved histidines alters the tertiary structure and nanomechanics of consensus ankyrin repeats, J. Biol. Chem. 287 (23) (2012) 19115–19121.

[180] M. Sotomayor, K. Schulten, Single-molecule experiments in vitro and in silico, Science 316 (5828) (2007) 1144–1148.
[181] T. Verdorfer, R.C. Bernardi, A. Meinhold, W. Ott, Z. Luthey-Schulten, M.A. Nash, H.E. Gaub, Combining in vitro and in silico single-molecule force spectroscopy to characterize and tune cellulosomal scaffoldin mechanics, J. Am. Chem. Soc. 139 (49) (2017) 17841–17852.
[182] D.E. Makarov, A theoretical model for the mechanical unfolding of repeat proteins, Biophys. J. 96 (6) (2009) 2160–2167.
[183] G.M. Nam, D.E. Makarov, Extracting intrinsic dynamic parameters of biomolecular folding from single-molecule force spectroscopy experiments, Protein Sci. 25 (1) (2016) 123–134.
[184] R.B. Best, S.B. Fowler, J.L. Toca Herrera, A. Steward, E. Paci, J. Clarke, Mechanical unfolding of a titin Ig domain: Structure of transition state revealed by combining atomic force microscopy, protein engineering and molecular dynamics simulations, J. Mol. Biol. 330 (4) (2003) 867–877.
[185] R.B. Best, G. Hummer, Protein folding kinetics under force from molecular simulation, J. Am. Chem. Soc. 130 (12) (2008) 3706–3707.
[186] C. Hyeon, D. Thirumalai, Measuring the energy landscape roughness and the transition state location of biomolecules using single molecule mechanical unfolding experiments, J. Phys.-Condensed Matter 19 (11) (2007) 113101, https://doi.org/10.1088/0953-8984/19/11/113101.
[187] M.S. Li, C.-K. Hu, D.K. Klimov, D. Thirumalai, Multiple stepwise refolding of immunoglobulin domain I27 upon force quench depends on initial conditions, PNAS 103 (1) (2006) 93–98.
[188] D. Thirumalai, Z. Liu, E.P. O'Brien, G. Reddy, Protein folding: from theory to practice, Curr. Opin. Struct. Biol. 23 (1) (2013) 22–29.
[189] B. Isralewitz, M.u. Gao, K. Schulten, Steered molecular dynamics and mechanical functions of proteins, Curr. Opin. Struct. Biol. 11 (2) (2001) 224–230.
[190] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, K. Schulten, Molecular dynamics study of unbinding of the avidin-biotin complex, Biophys. J. 72 (4) (1997) 1568–1581.
[191] S. Izrailev S. Stepaniants K. Schulten, Applications of steered molecular dynamics to protein-ligand/membrane binding. Biophys. J., 1998. 74(2): p. A177–A177.
[192] E. Evans, Probing the relation between force - Lifetime - and chemistry in single molecular bonds, Annu. Rev. Biophys. Biomol. Struct. 30 (1) (2001) 105–128.
[193] E. Evans, K. Ritchie, Dynamic strength of molecular adhesion bonds, Biophys. J . 72 (4) (1997) 1541–1555.
[194] C. Clementi, H. Nymeyer, J.N. Onuchic, Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? an investigation for small globular proteins, J. Mol. Biol. 298 (5) (2000) 937–953.
[195] B. Hess, et al., GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, J. Chem. Theory Comput. 4 (3) (2008) 435–447. [196] V. Bennett, D.N. Lorenzo, in: Chapter One - Spectrin- and Ankyrin-Based Membrane Domains and the Evolution of Vertebrates, in Current Topics in Membranes, Academic Press, 2013, pp. 1–37.
[197] V. Bennett, D.N. Lorenzo, An adaptable spectrin/ankyrin-based mechanism for long-range organization of plasma membranes in vertebrate tissues, in: V. Bennett (Ed.), Dynamic Plasma Membranes: Portals between Cells and Physiology, Elsevier Academic Press Inc, San Diego, 2016, pp. 143–184.
[198] S.K. Wetzel, G. Settanni, M. Kenig, H.K. Binz, A. Plückthun, Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins, J. Mol. Biol. 376 (1) (2008) 241–257.
[199] M. Rief, J. Pascual, M. Saraste, H.E. Gaub, Single molecule force spectroscopy of spectrin repeats: Low unfolding forces in helix bundles, J. Mol. Biol. 286 (2) (1999) 553–561.
[200] G. Lee, K. Abdi, Y. Jiang, P. Michaely, V. Bennett, P.E. Marszalek, Nanospring behaviour of ankyrin repeats, Nature 440 (7081) (2006) 246–249.
[201] A.N. Lupas, M. Gruber, in: The Structure of α-Helical Coiled Coils, in Advances in Protein Chemistry, Academic Press, 2005, pp. 37–38.
[202] T. Bornschlogl, M. Rief, Single-molecule dynamics of mechanical coiled-coil¨ unzipping, Langmuir 24 (4) (2008) 1338–1342.
[203] T. Bornschlogl, M. Rief, Single molecule unzipping of coiled coils: sequence¨ resolved stability profiles, Phys. Rev. Lett. 96 (11) (2006), 118102.
[204] M. Kim, et al., Fast and forceful refolding of stretched alpha helical solenoid proteins, Biophys. J. 98 (12) (2010) 3086–3092.
[205] Z.N. Scholl, et al., Single-molecule force spectroscopy reveals the calcium dependence of the alternative conformations in the native state of a – crystallin protein, J. Biol. Chem. 291 (35) (2016) 18263–18275.