



UNIVERSITY OF LEEDS

This is a repository copy of *The physics of pulling polyproteins: a review of single molecule force spectroscopy using the AFM to study protein unfolding.*

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/101530/>

Version: Accepted Version

Article:

Hughes, ML and Dougan, L (2016) The physics of pulling polyproteins: a review of single molecule force spectroscopy using the AFM to study protein unfolding. Reports on Progress in Physics, 79 (7). 076601. ISSN 0034-4885

<https://doi.org/10.1088/0034-4885/79/7/076601>

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

The physics of pulling polyproteins: a review of single molecule force spectroscopy using the AFM to study protein unfolding

Megan Hughes and Lorna Dougan

*School of Physics and Astronomy
University of Leeds,
United Kingdom, LS2 9JT*

E-mail: L.Dougan@leeds.ac.uk

Contents

1	Abstract	2
2	Introduction	2
3	Protein structure and stability	3
3.1	Introduction to protein structure	3
3.2	Classes of proteins	5
3.3	Protein thermodynamic and kinetic stability	7
4	Relevant forces for proteins	9
5	Single-molecule force spectroscopy of polyproteins	13
5.1	Fundamental principles of the atomic force microscope	13
5.2	Calibration of the cantilever	14
5.3	Polyproteins	14
5.4	Force-extension protein unfolding	16
5.5	Analysis of force-extension data	19
6	Stretching proteins: polymer elasticity models	20
6.1	The freely jointed chain model	20
6.2	The worm-like chain model	21
6.3	Further developments in polymer elasticity models	23
7	Probing the effect of force on the kinetic and thermodynamic protein unfolding process	25
7.1	Dependence of the unfolding force on the loading rate	25
7.2	Kinetic parameters of protein mechanical unfolding	27
7.2.1	The Bell-Evans-Richie Model	27
7.2.2	Monte Carlo simulations and the Bell-Evans-Richie model	28
7.2.3	Friddle-de-Yoreo model	29
7.2.4	Dudko-Hummer-Szabo model	30
7.3	Thermodynamic parameters of protein mechanical unfolding	32
7.3.1	Jarzynski equality	33
7.3.2	Crooks fluctuation theorem	33

8 Protein mechanical stability	34
8.1 Effect of unfolding history on measured protein mechanical stability	34
8.2 Relationship between protein structure and mechanical stability	34
8.3 Importance of the direction of the applied force	34
8.4 Mechanical clamps motifs in proteins	38
8.5 Protein engineering to tune mechanical stability	38
8.6 Role of solvent interactions on protein mechanical stability	41
8.7 Temperature softening of protein mechanical properties	41
8.8 Measuring energy landscape roughness of proteins using mechanical unfolding experiments . .	42
8.9 Mechanical Phi-value analysis probes mechanical unfolding transition state	42
9 Towards design rules for predicting the mechanical properties of proteins	42
9.1 Predicting the mechanical stability and malleability of proteins	42
9.2 Predicting the critical rupture force of proteins	45
9.3 Propagation of forces in proteins	45
10 Future Directions	47
10.1 Lessons from extremophiles: SMFS studies of extreme-loving proteins	47
10.2 Folded protein-based biomaterials	48
11 Conclusions	48
12 Acknowledgements	48

1 Abstract

One of the most exciting developments in the field of biological physics in recent years is the ability to manipulate single molecules and probe their properties and function. Since its emergence over two decades ago, single molecule force spectroscopy has become a powerful tool to explore the response of biological molecules, including proteins, DNA, RNA and their complexes, to the application of an applied force. The force versus extension response of molecules can provide valuable insight into its mechanical stability, as well as details of the underlying energy landscape. In this review we will introduce the technique of single molecule force spectroscopy using the atomic force microscope (AFM), with particular focus on its application to study proteins. We will review the models which have been developed and employed to extract information from single molecule force spectroscopy experiments. Finally, we will end with a discussion of future directions in this field.

2 Introduction

Proteins are biological polymers made up of monomeric units called amino acids. They are “bionanomachines” and are responsible for a vast array of biological functions that occur *in vivo*. Acting in isolation, or as part of larger and often complex machinery, they perform their function through structural changes. These structural changes involve breaking and formation of their intra- or intermolecular interactions and a delicate interplay of interactions with the surrounding solvent environment. The magnitudes of the energies involved in protein conformational changes are similar to thermal energy, resulting in stochastic and noisy behaviour in biological systems. Despite this complexity, proteins can be understood qualitatively using simplified rules and models such as entropic elasticity, as well as through an understanding of important non-covalent interactions such as van der Waals forces, electrostatic interactions, hydrogen bonding and hydrophobic interactions. Most protein molecules are synthesised as flexible chains, that then fold into a well-defined three dimensional structure, the native conformation. Protein folding is therefore an impressive example of self-assembly. Proteins use mechanical forces in different cellular processes, ranging from replication, transcription, translation and protein degradation, to cell adhesion and transport. Considerable

efforts are directed towards understanding how proteins move and change shape in response to the mechanical and thermal forces exerted on them. Understanding the structure and dynamics of these biological bionanomachines remains one of the major frontiers in science.

In the last two decades, the field of single molecule force spectroscopy (SMFS) has emerged as a powerful approach to investigate the mechanical processes which involve proteins at the single molecule level. SMFS enables single protein molecules to be unfolded over a well-defined reaction co-ordinate. Proteins can be mechanically unfolded by a number of experimental techniques [1] including optical tweezers [2], magnetic tweezers [3, 4] and the atomic force microscope (AFM)[5–8, 8, 9]. For each approach, the protein is typically tethered at one end to a surface or tip leaving the other end free to interact with either the tip, another molecule or a surface. The force response of the protein is then monitored by either unfolding the protein at constant velocity (force-extension mode)[8], or under a constant force (force-clamp mode)[10]. The aim of this review is to provide the non-specialist with a comprehensive introduction to the technique and models employed for SMFS. It will focus on the underlying physics which has enabled the experimental methodology and the interpretation of the data. We begin with an introduction to the structure and stability of proteins. Several manipulation techniques can now be employed to study the mechanical properties of single proteins. Here we focus on SMFS of proteins using the AFM in force-extension mode, one of the manipulation techniques most extensively used for the study of the mechanical properties of proteins. Whilst this review focusses on the AFM, we would like to point our readers to several other available reviews outlining the vast contributions of optical tweezers [11–13] and magnetic tweezers [14–16] to SMFS. This review will start by introducing the biological molecules proteins but will not include an introduction to DNA - another extensively studied biological molecule in SMFS. We would like to point our readers to a recent review for an introduction to DNA molecules and the results of studying these molecules by SMFS [17]. Our review includes an introduction to the main components of the AFM for SMFS. We describe the main mode of operation of this instrument for the mechanical manipulation of proteins and the polymer physics models employed to reveal information on the underlying force-extension behaviour of polypeptide chains. Much has been learned about the role of non-covalent interactions in determining protein mechanical stability. We provide a detailed review of the key findings in this area, and highlight the seminal work which has provided step changes in our understanding. In parallel with advances in experimental techniques, theoretical models have been developed to interpret experimental data from single molecule protein unfolding. We describe a number of these models and present examples of their application to experimental data. We discuss efforts to predict the mechanical stability of proteins which rely on empirical models as well as approaches appropriated from materials science. Finally we discuss emerging new directions in the field of SMFS.

3 Protein structure and stability

3.1 Introduction to protein structure

Proteins are polymers (also known as polypeptides) made up of monomer subunits called amino acids which are covalently linked together via peptide bonds [18]. An amino acid is an organic molecule consisting of a central carbon atom covalently bound to a hydrogen, a carboxylic acid group, an amino group and a side chain. The typical chemical structure of an amino acid is shown in Figure 1 where R denotes the side chain. There are 20 naturally occurring amino acids each with a unique side chain with different physical and chemical properties (Figure 1). The order of amino acids in the polypeptide chain is known as the sequence of the protein and forms the primary structure of the protein as shown in Figure 2. The primary sequence determines how the polypeptide chain folds, governed by the type and location of each amino acid and their ability to interact with other parts of the chain and the surrounding solvent environment (Figure 2).

The most common secondary structural arrangements or motifs a protein adopts are known as β -sheets, formed by parallel or anti-parallel β -strands, and α -helices (Figure 2), the latter of which can have different classes which define the pitch of the helix. The existence of these structures was predicted as early as 1951 by Linus Pauling, who received a Nobel prize for his contribution [20, 21]. The protein’s tertiary structure refers to the geometric shape of the molecule (Figure 2C) and can be classed in three main categories: all β structure; all α structure and α/β structure. The quaternary structure of a protein refers to interactions between multiple polypeptide chains or subunits. When a protein folds, it tests multiple 3D structures before adopting the most favourable conformation. This unique structure is held together by many non-covalent

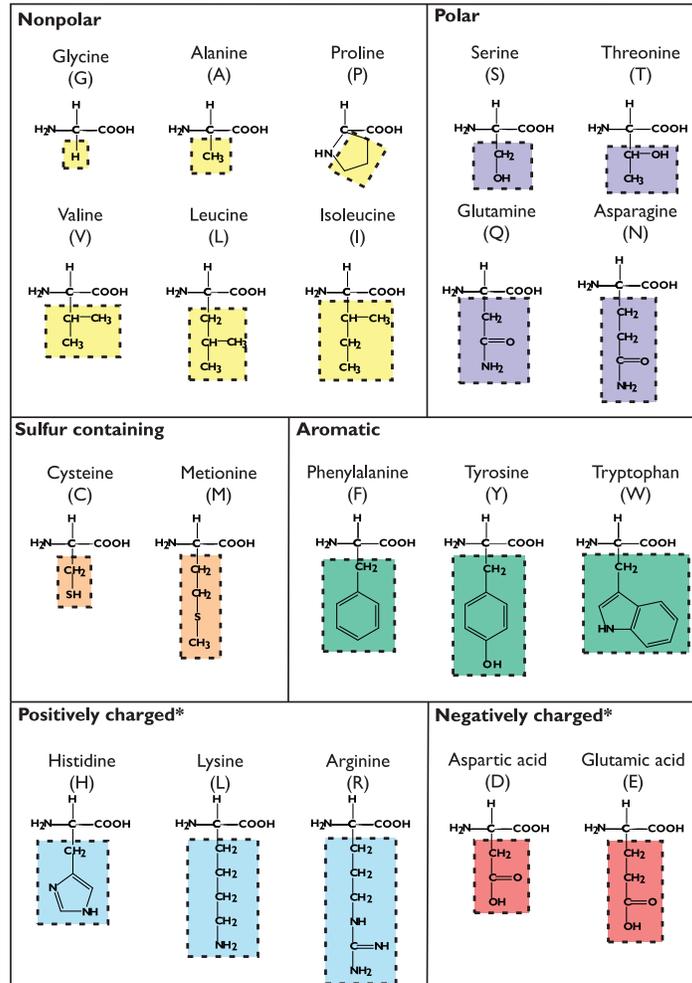
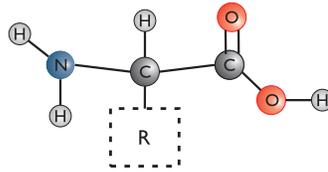


Figure 1: Proteins are made up of subunits called amino acids. An amino acid is made up of a central carbon atom, known as the α -carbon, covalently bound to a hydrogen atom, an amino group (NH_2), a carboxyl group (COOH) and a side chain group (R group). The side chains of the 20 standard amino acids are shown in this figure, where the amino acids are grouped according to the properties of their side chains [19]. Nonpolar, aliphatic amino acids (yellow) have hydrocarbon side chains and are typically hydrophobic. The aliphatic polar uncharged amino acids (purple) contain an amino or hydroxyl group and can form hydrogen bonds with atoms in other polar amino acids or water molecules. Aromatic amino acids (green) contain an aromatic ring and can be nonpolar or polar. The sulfur containing amino acids (orange) are named cysteine and methionine. The side chain of methionine is hydrophobic. The side chain of cysteine can form covalent disulphide bonds with other cysteine residues due to the sulfur-hydryl (SH) group found on its side-chain. At neutral pH the charged amino acids can be either charged negative, (acidic, red) or positive (basic, blue).

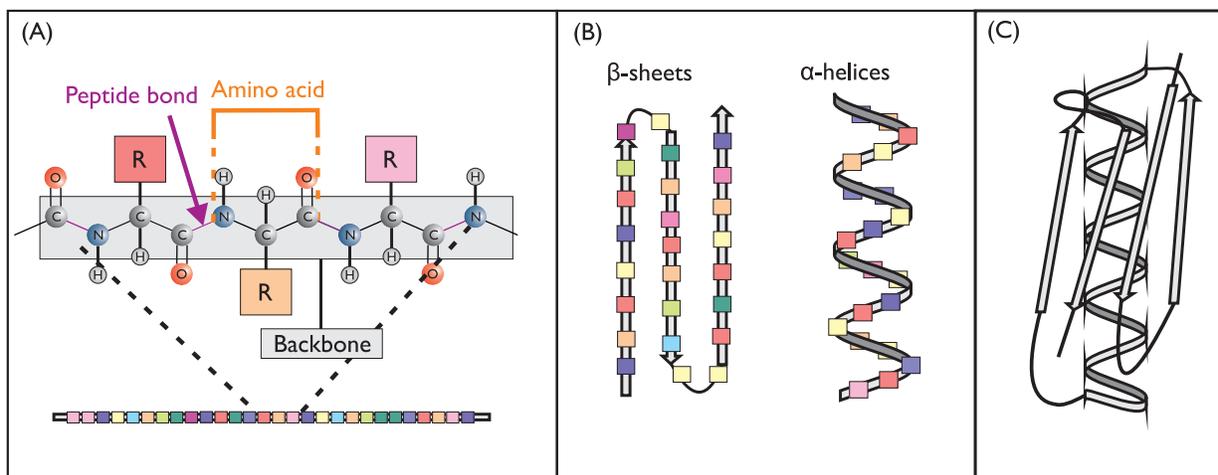


Figure 2: Schematic showing the primary, secondary and tertiary structures of proteins. (A) The order of amino acids in a protein is known as the primary structure. The different coloured squares refer to different amino acid side chains. (B) Based on the primary structure of the protein, this chain can fold into different shapes known as the secondary structure of the protein. (C) The protein's tertiary structure refers to the geometric shape of the molecule, shown here representing the structure of protein L (PDB ID:1HZ6).

interactions between the polypeptide chain and with the surrounding environment. Sir John Kendrew and Max Perutz solved the first protein structure using X-ray diffraction in the 1950s, earning them a shared Nobel prize in Chemistry in 1962 'for their studies of the structures of globular proteins'. While the most common structural motifs, such as α -helices and β -sheets, had already been predicted by Pauling it was only with the determination of first protein structures that they could be confirmed. The protein database (PDB) now contains 3D structural data on more than 100 000 biological molecules.

3.2 Classes of proteins

There are three main classes of proteins globular, membrane and fibrous. These are classified by the properties of their environment, the interactions stabilising the proteins and their architecture. Schematics of the three classes are shown in Figure 3. This review will focus on globular proteins. Membrane proteins and their interactions with ligands have been extensively studied by SMFS techniques [22, 23]. We would like to point our readers to specific reviews on the single molecule force spectroscopy of membrane proteins for more detailed information about the recent contributions SMFS has had to this class of protein [24–28].

Globular proteins

This class of protein is water soluble and as a result is the most straightforward to study. The proteins often form a globule shape, which consists of different secondary structural elements. These secondary structural elements typically form to shield hydrophobic amino acids from contact with the water environment surrounding them. There are three subclasses of this protein type according to their structure: all α , all β and α/β .

Membrane proteins

Cells are surrounded by a membrane which separates internal components of the cell and external components of the environment. Membranes are formed by a combination of lipids and proteins. Lipids are small biological molecules that have a hydrophilic (head) and hydrophobic (tail) component. This causes them to arrange themselves into layers separating the hydrophobic tails of the lipids from contact with water (Figure

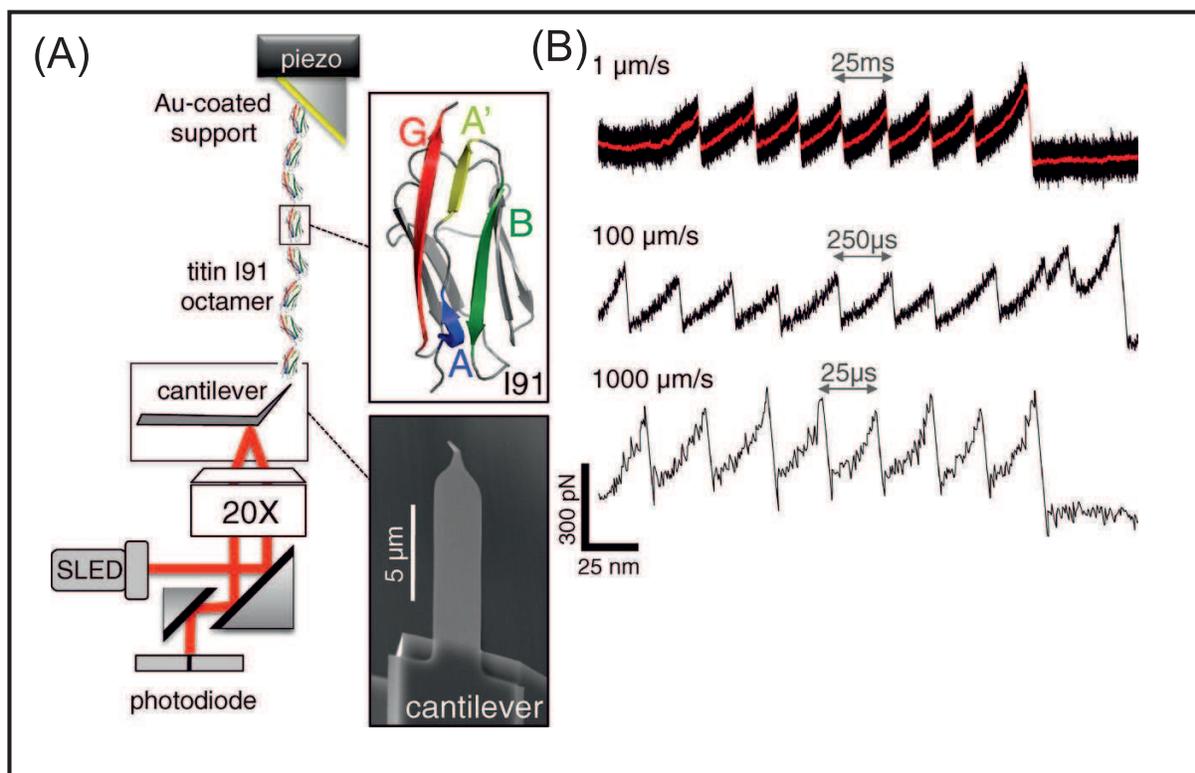


Figure 3: Schematic showing the difference classes of proteins where waters are shown in light blue. (A) Globular proteins: these proteins are water soluble, often hydrophobic amino acids are found to be located in the center of the protein. This is illustrated by the yellow core in the schematic. (B) Membrane proteins: Cells are surrounded by a thin membrane composed of lipids and proteins. Lipids are naturally occurring molecules, such as fats, with a polar region and hydrophobic region. In this diagram the hydrophobic regions are illustrated by the black tails attached to a polar head regions (grey spheres). These organise themselves to minimise the interaction between the hydrophobic tails and water molecules causing them to form a bilayer, where the heads provide two hydrophilic surfaces and a hydrophobic (yellow) region in between. Proteins can be found embedded in this membrane. In this example the protein traverses through the membrane and the hydrophilic regions of the protein is exposed to the water molecules either side of the membrane. (C) An example of a fibrous protein is collagen, a non-soluble, structural protein. Collagen is a fibril which is a superhelix formed by three interlinking polypeptide chains.

3), this results in an ordered structure with a defined hydrophobic region. Membrane proteins have many functions. They are responsible for transport of various molecules across the membrane layers, signalling and sensing, often through mechanical triggers or by mechanical changes to the protein. These proteins can be fully embedded into the lipid layer (Figure 3 (B)) where the regions of the protein within the membrane are highly ordered with regular secondary structure, whereas regions extruding from the membrane are often hydrophilic and often disordered. Membrane proteins can contain multiple α -helices that traverse multiple times through the membrane, or multiple β -sheets that can form a barrel shape to provide a channel through the membrane. The proteins can also just traverse through the membrane once, these are often single α -helices, or can also be anchored to one side of the membrane layer.

Fibrous proteins

These long chain proteins serve as structural proteins to maintain the structure of cells and tissues. These proteins are often found to have regular repeating regions of amino acids which cause regions of highly regular structure within the protein. Multiple long polypeptide chains can intertwine and link together to form the 3D structure of the protein and giving it special bulk properties. Fibrous proteins often aggregate (many associate) to form higher-ordered filaments and fibrils. This enables them to perform many functions such as providing structural support to cells and tissues. These structures can be further separated into three subcategories:

- β -structural - made up of β structures. One example is silk fibroin, a component of silk fiber, which is formed of many β sheets.
- α -structural - these can be coiled coil helices which are two or more helical polypeptide chains that wind together to form a superhelix. Many supercoiled structures can often associate to form fibrils.
- Collagen - these are non soluble fibrils formed by superhelices made from three intertwined polypeptide chains that solely interact between each other (inter) rather than within the single polypeptide (intra). These proteins are found within the bone matrix - changes within the primary structure of these proteins can cause diseases such as brittle bone disease.

Often large fibrous proteins can be made up of smaller globular proteins, for example the large fibrous muscle protein titin is made up of combination of regions containing globular proteins or fibrous proteins.

3.3 Protein thermodynamic and kinetic stability

The native conformation of a protein represents a minimum of its thermodynamic Gibbs free energy, G . For a system at constant pressure, and constant temperature, measuring changes in the Gibbs free energy, ΔG , for an alteration in the state of the system can be measured. The ΔG is a combination of the change in enthalpy ΔH and change in entropy of the system ΔS , ($\Delta G = \Delta H - T\Delta S$) and is used to determine whether a reaction is favourable or unfavourable. In the context of protein unfolding, the thermodynamic stability of the protein is defined as the difference in free energy of the unfolded protein and the folded protein (Figure 4A) and is the work required to transfer a body from one state to another whilst the body exchanges heat with the environment [18]. This simple two-state energy landscape (Figure 4A) describes the process of converting from the folded (F) and unfolded (U) protein where the probability of this process is governed by the rate of unfolding (k_U) and the rate of folding (k_F). The difference in the free energy of these two states, ΔG_{U}^0 determines the thermodynamic stability of the protein. A barrier to unfolding separates the two states of the protein. This barrier is known as the transition state (TS) of the protein and the height of this barrier relative to the F state is ΔG_{TS}^0 . Many proteins have more complex landscapes [29], where proteins fold via stable intermediates (Figure 4B) due to multiple barriers along the reaction coordinate probed, which might include the number of native contact or a length. Statistical mechanics-based models postulate that protein molecules traverse an energy landscape during folding, and that protein folding pathways more closely resemble funnels than random diffusion in configuration space [29, 30]. A folding funnel is a plot of the free energy and the conformational space. An individual folding trajectory is envisaged for each polypeptide chain traversing down the folding funnel (Figure 4C) [29, 30] where the proteins are biased to the folded state and therefore many intermediates and folding pathways exist causing the energy landscape to be rough. The

folding energy landscape of membrane proteins have further complexities to do with the interactions between the membrane and the environment, the complexities of the energy landscapes of membrane proteins are discussed in this review [31].

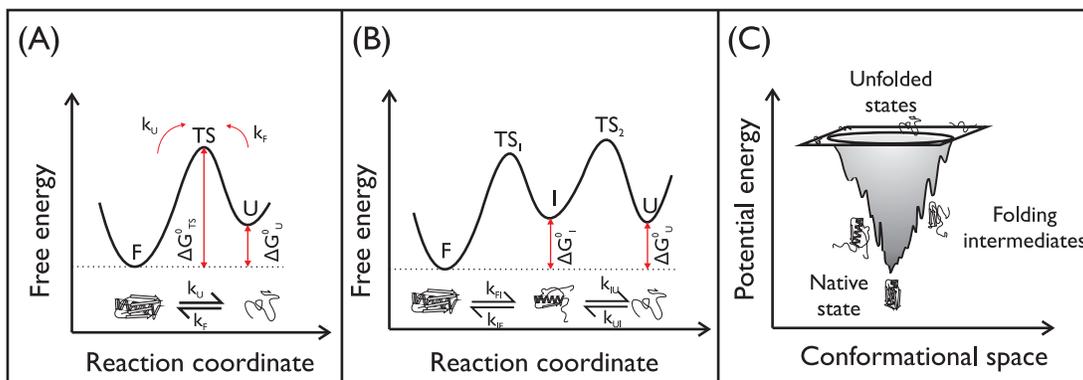


Figure 4: (A) Schematic of a two-state energy landscape of a protein, where the unfolding (U) and folded (F) states of a protein are separated by a single barrier known as the transition state (TS) of the protein. The difference in the free energy of the U and F states, ΔG_U^0 determines the thermodynamic stability of the protein. The height of the TS relative to the F state is ΔG_{TS}^0 and is the energy required to overcome the barrier to unfolding. Bottom: The forward arrow indicates the process of going from the F state to the U state, the backwards arrow indicates the process of the U state going to the F state. The probability of these processes are quantified by the rate of unfolding (k_U) and the rate of folding (k_F) respectively. (B) A three-state energy landscape containing three energy minima in the profile. The minima correspond to a folded (F) protein, a stable intermediate (I) state and the unfolded (U) protein. Each minimum is separated by a barrier. The first energy barrier, TS₁, separates the I and F states. The second barrier separating the I and U protein states is, TS₂. The probability of the protein going from the F state to the I state, and the I state to the F state is governed by k_{FI} and k_{IF} respectively. The probability of the protein going from the I state to the U state and the U state to the I state is governed by k_{IU} and k_{UI} respectively. (C) Schematic of the energy landscape profile of a protein in 3D conformational space. A large amount of conformational freedom is available to the unfolded state of the protein in the 3D conformational space. This conformational freedom decreases as the protein folds. This decrease is shown on the schematic as a transition from white to black. The corresponding energy profile of the protein represents a funnel with multiple transition states between the unfolded and folded states. The protein tends towards the folded, native state of the protein as the free energy of the system decreases. This native state corresponds to the global minimum in the schematic.

For protein folding to be spontaneous, the free energy between the unfolded and folded state, ΔG_U^0 , must be negative [18]. When a protein is unfolded there are a vast number of conformations available to the polypeptide chain. When a protein folds it has a fewer conformations available to it [32], resulting in a decrease in the entropy of the system. This entropy is known as the conformational entropy of the system, ΔS_{conf} and contributes unfavourably to ΔG_U^0 [32, 33]. The favourable contribution to ΔH arises from non-covalent interactions between parts of the polypeptide chain. These interactions include: hydrogen bonds, Van der Waals interactions, salt bridges and ion pairs [18, 33–35]. The importance of hydrogen bonds (HBs) in the structure of proteins was first made apparent by Pauling *et al.*, who predicted the existence of α -helical and β -sheets in the secondary structures of proteins (Figure 5) based solely on the knowledge of the peptide bond and the existence of HBs [20, 36, 37]. Proteins contain many atoms that are capable of forming HBs and these interactions underpin the secondary structure of the protein (Figure 5D).

The tertiary structure of proteins is “glued” together by hydrophobic interactions, which refer to the association of nonpolar solutes to minimise their interaction with water [38]. The importance of hydrophobic interactions in biological systems was noted by Kauzmann in 1959 [39] who predicted ‘hydrophobic interac-

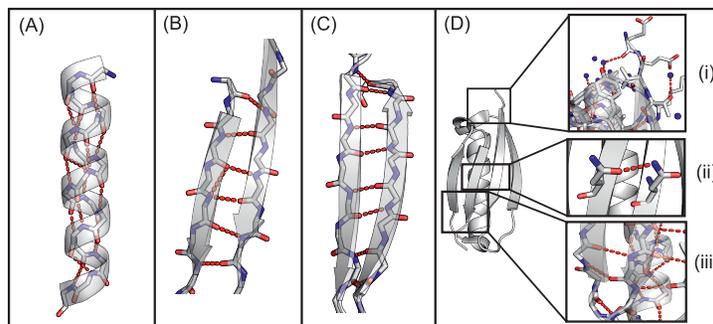


Figure 5: Schematic indicating the hydrogen bonds within the protein backbone formed between hydrogen atoms in the amide group and oxygen atoms in the carbonyl groups. These hydrogen bonds are illustrated by dashed red lines. The oxygen of the carbonyl is coloured red, the nitrogen of the amide group is coloured blue and the carbon backbone is coloured grey. These hydrogen bonds underpin the structural stability of (A) α -helices and (B) and (C) β -strands. β -strands can be arranged in parallel (B) or anti-parallel (C) geometry based on the chain direction. (D) Hydrogen bonds underpin the secondary structure of the protein.

tions' to be a major driving force in protein folding. The 'hydrophobic effect' was introduced by Tanford in 1980 to describe this phenomenon. To help elucidate the contribution of hydrophobic interactions to protein stability, a number of studies have modelled the formation of the hydrophobic core of the protein as the transference of nonpolar side chains from water to a nonpolar environment [38, 40]. By studying the $\Delta G_{\text{water-nonpolar}}$, the change in free energy when a solute is transferred from water to a nonpolar environment, for amino acids, or chemical derivative, hydrophobicity scales for the 20 naturally occurring amino acids have been determined [41–44]. This is illustrated in Figure 6 which shows examples of the hydrophobicity rankings of the amino acids determined by different groups and methods. As well as non-covalent interactions in the protein, the solvent environment must be considered. Water forms a hydration layer around proteins which is dynamic and exchanges with bulk water due to thermal fluctuations on the surface of the protein [45, 45–50]. The structure of the protein can be maintained by this hydration layer [51] and conformational changes are influenced by hydration dynamics [52]. The hydrogen bonding network within the water must therefore be dynamic and flexible to facilitate structural changes associated with the function of the protein [53].

Hydrophobic interactions and hydrogen bonds are thought to be two of the most important contributions to protein stability [33]. For example, Figure 7 shows the predicted contributions of different interactions to the protein Ribonuclease Sa, which contains 96 amino acids. The destabilising contribution of $T\Delta S_{\text{conf}}$ towards protein folding is predicted to be ~ 700 kJ/mol. This is counteracted by the combination of hydrogen bonds (~ -350 kJ/mol) and hydrophobic interactions (~ -400 kJ/mol), resulting in a small, negative ΔG_{U}^0 .

4 Relevant forces for proteins

Given the important role of non-covalent interactions in protein stability and dynamics, what is the magnitude of biologically relevant forces that affect proteins? Proteins experience thermal agitation due to Brownian fluctuations of energy $k_{\text{B}}T = 4 \times 10^{-21} \text{J} = 0.6 \text{ kcal mol}^{-1} = 4.1 \text{ pNnm}$ at room temperature. The forces associated with different length-scales corresponding to different intermolecular interactions and bonds are shown in Figure 8. Given that proteins are subjected to thermal forces, the number of possible configurations (entropy) is at its maximum when a protein forms a random coil or is denatured. This number reduces significantly when the protein forms secondary and tertiary structures. Upon stretching, the molecular entropy is reduced so that at full extension there is only one backbone configuration, a fully elongated protein. Since the typical energies involved are of order $k_{\text{B}}T$, and the typical lengths are of the order of nanometres, entropic forces are on the order of piconewtons, depending on the pulling speed. Stretching a protein in the low force regime, to overcome entropic forces, can be achieved using SMFS. Non-covalent

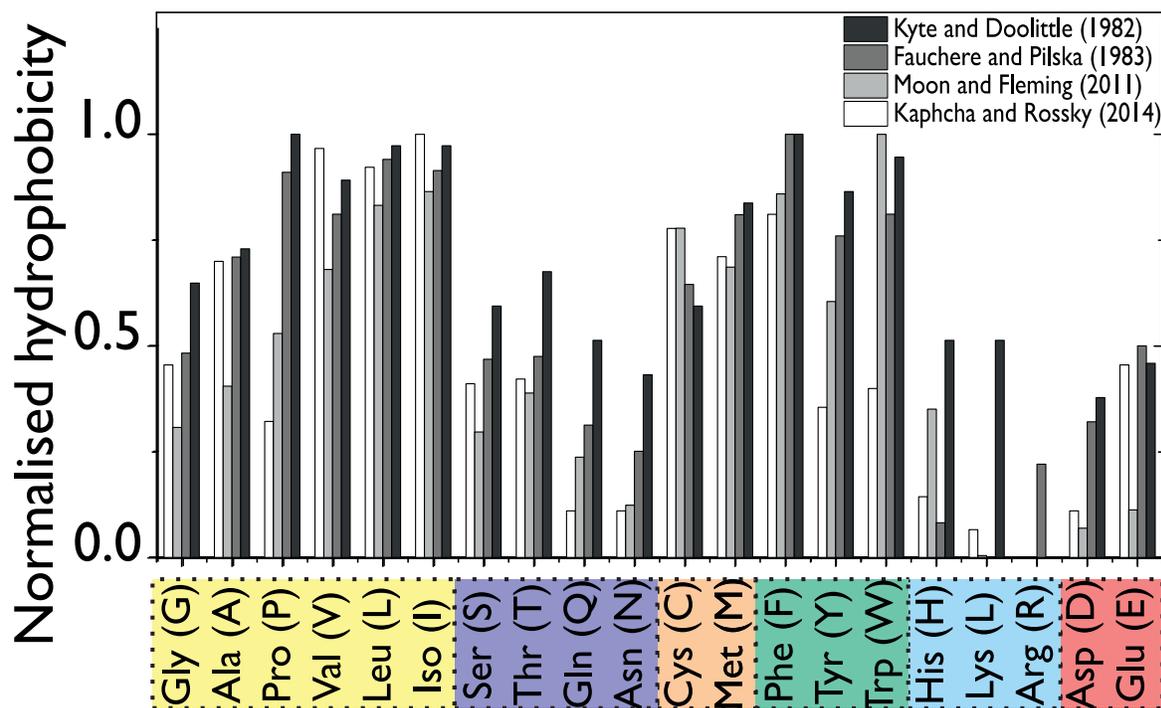


Figure 6: To determine the hydropathy scales or hydrophobicity of amino acids, studies have used nonpolar environments such as bilayers [54], liquid alkanes or alcohols [43, 55], which are slightly polar. The tightly packed hydrophobic core is slightly polar due to the enhanced van der Waals interactions between amino acid side chains [56]. Many scales have therefore been derived because it is challenging to determine which environment best mimics that of the protein core [40]. As a result there are often discrepancies between the available hydrophobic scales. This graph shows the normalised hydrophobic scales from four groups for the 20 naturally occurring amino acids where 1 corresponds to the most hydrophobic amino acid and 0 corresponds to the most hydrophilic amino acid. The Kyte-Doolittle scale (black) is taken from ref. [42], this scale was determined from the position of an amino acid within the tertiary structure of a number of proteins. The Fauchere-Pilska scale is taken from [18] (dark grey) and these values are determined from the transfer of amino acids from water to alcohol. The Moon-Fleming scale from ref. [54] (lighter grey) is determined from transferring amino acids from water to bilayers. Finally the Kaphcha and Rosicky scale [41] is derived from simple coarse grain computational modelling.

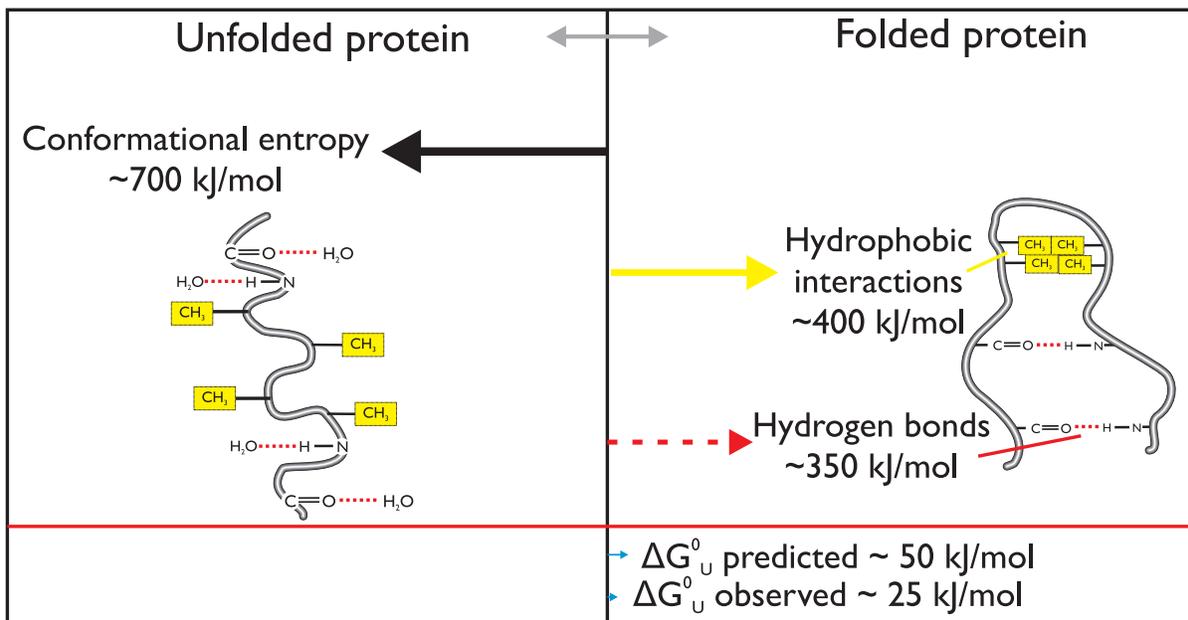


Figure 7: Schematic of a protein folding process illustrating the formation of hydrogen bonds and hydrophobic interactions and the negative and positive contributions to free energy of the folded protein Ribonuclease Sa (at 25 °C, pH 7). The corresponding free energy values are taken from experimental studies [32]. The conformational entropy contributes unfavourably to the free energy of (destabilises) the folded protein. Hydrogen bonds and hydrophobic interactions contribute favourably to the stability of the folded protein. Figure and values adapted from Figure 2 (a and b) Pace, C. N., Grimsley, G. R., Scholtz, J. M. and Shaw, K. L. 2014. Protein Stability. eLS. © 1999-2015 John Wiley Sons, Inc. All Rights Reserved.

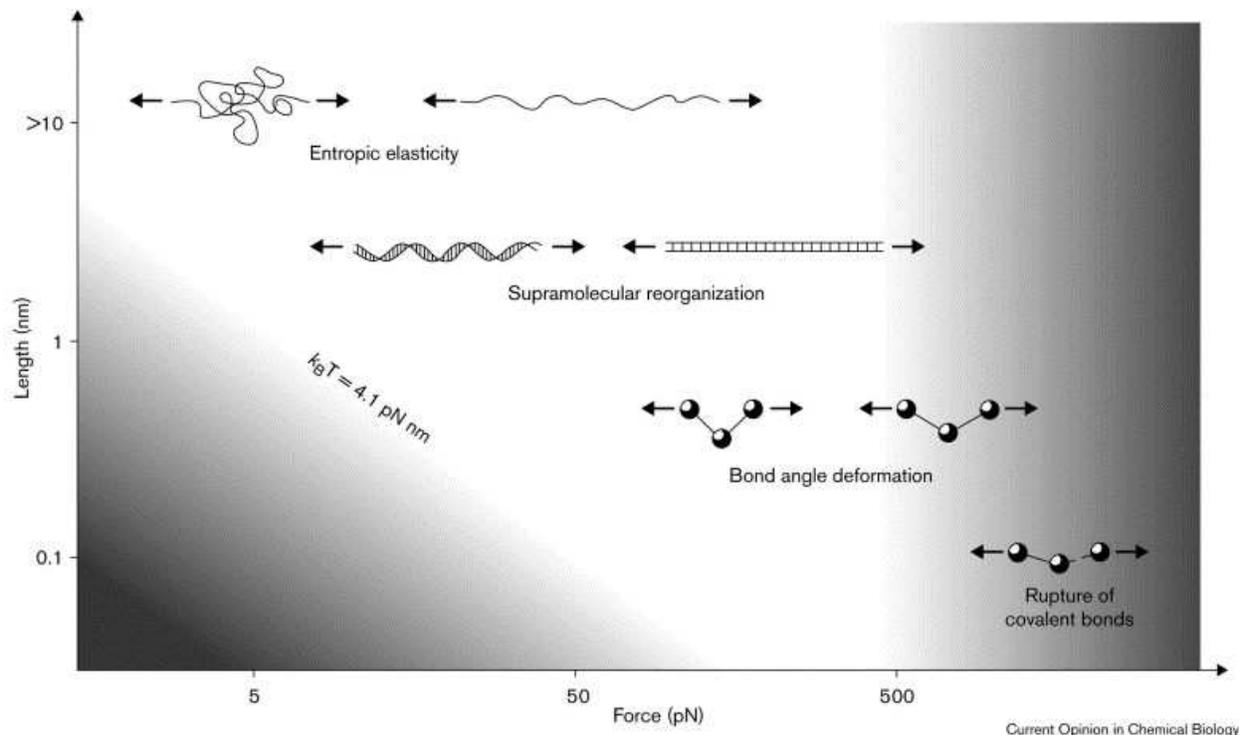


Figure 8: Figure illustrating the lengthscales, and corresponding forces associated with bond breakage, for different intermolecular interactions. The figure highlights the forces available to SMFS by a white region of the plot. Inaccessible regions are shaded, these regions correspond to the limit of thermal stability of molecular structures (low forces, left) and the breakage of covalent bonds (high forces, right). Reprinted from [7] *Current Opinion in Chemical Biology*, 4, H. Clausen-Schaumann, M. Seitz, R. Krautbauer, H. E. Gaub, Force spectroscopy with single bio-molecules, Pages 524–530, Copyright (2000), with permission from Elsevier.

interactions are very important in proteins. Although these bonds are weaker than covalent bonds, they are numerous enough to make proteins strong. Unfolding a protein involves modifications of the molecular structure on a nanometre lengthscale, requiring breakage and reformation of many van der Waals interactions, hydrogen bonds and electrostatic interactions. The forces involved in breaking these combinations of bonds are typically in the range 100-300 pN, when measured at loading rates typical of an AFM experiment (200-1000 nm/s). The strongest forces encountered at the molecular scale are those required to break covalent bonds. SMFS techniques have determined that the force required to break a covalent bond is of the order of 1600 pN [57].

There are now many wonderful examples of the role of force in biological systems, and many of these have been probed using SMFS methods. Force acts as a denaturant in SMFS experiments. As the force is applied across the protein domains the activation barrier for unfolding is lowered, increasing the probability that the protein will unfold. Studies have helped identify the role of mechanical unfolding forces in protein degradation [58–60]. They have uncovered details of force generation in motor proteins [61, 62] and the importance of plasticity of hydrogen bond networks in regulating the mechanochemistry of cell adhesion complexes [63], cell signalling (mechanosensors) [64], force generation [61, 62] and cell signalling [65, 66]. We point the reader to a recent review which discusses how force can be used to drive conformational changes in proteins, as well as modulate their stability and the affinity of their complexes[67].

5 Single-molecule force spectroscopy of polyproteins

5.1 Fundamental principles of the atomic force microscope

The AFM was invented in 1986 by Binnig, Quate and Gerber and the first commercially available AFM was introduced in 1989. The AFM is a high-resolution scanning probe microscopy instrument which allows mechanical manipulation of materials with atomic resolution. The most common use of the AFM is as a high-resolution imaging tool, where it is possible to measure the roughness of a sample surface. The AFM can also be used to probe and manipulate atoms and molecules in its SMFS mode of operation. The two essential components of the AFM are the cantilever, which acts as a flexible sensor, and a piezoelectric positioner, to provide accurate nanometre control of the sample position. The AFM can be used to measure forces with a sensitivity of 10 - 1000s of piconewtons (pN) and changes in length with nanometre resolution, while the timescale resolution is in the submillisecond range.

The cantilever is mounted onto a cantilever holder and the position of the cantilever with respect to the surface is controlled by a piezo-electric device. A focussed laser beam is reflected off the surface of the cantilever onto a photodetector. The deflections of the cantilever can be monitored by movement of the focussed laser spot on the photodiode. The four quadrant photodiode output (Figure 9) converts the incident light into voltage and then outputs the voltage difference, when the laser spot moves. The AFM transduces the forces exerted on its flexible cantilever by measuring the angular deviation of a laser beam. As the cantilever behaves as a Hookean spring and the force (F) is calculated as the product of the cantilever deflection in the z-axis (Δz_C) and its spring constant (k_c)

$$F = -k_c \Delta z_C \quad (1)$$

Any movement of the cantilever Δz_C , is converted into voltage differences ΔV by a split photodetector. At the start of every SMFS experiment the cantilever is calibrated. The k_c of the cantilever must be determined, as well as the optical lever sensitivity, which relates the photodiode output voltage to the cantilever deflection.

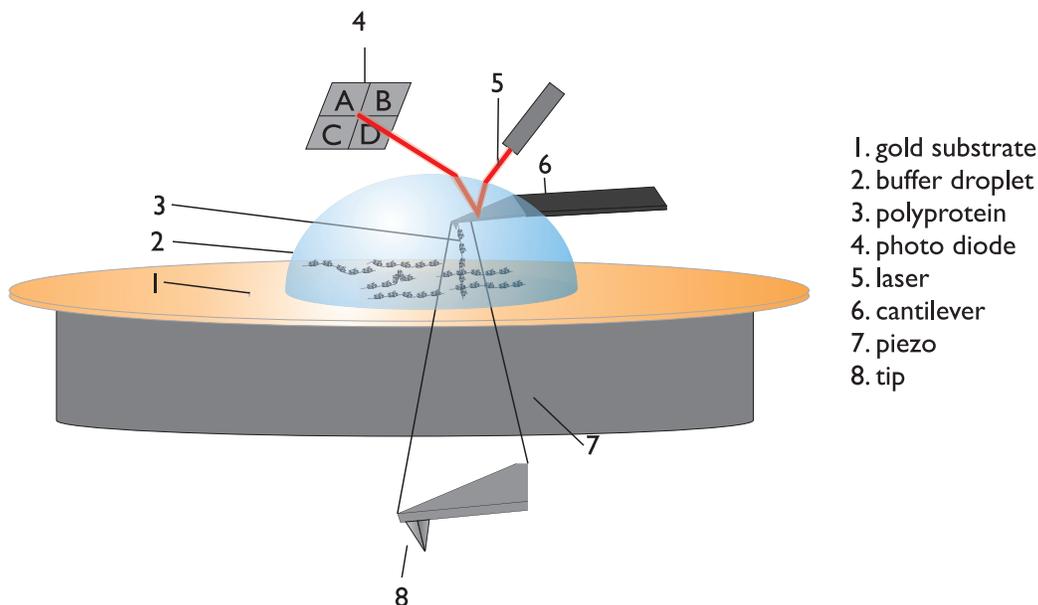


Figure 9: Schematic showing the main components of an AFM for SMFS. The protein is bound at one end to a substrate, in this example gold, which is covered in a suitable buffer. The gold substrate is placed on top of a piezoelectric device. A laser beam, reflected off of the cantilever tip into a photo-detector, is used to monitor the deflections of the cantilever.

5.2 Calibration of the cantilever

The ratio between the photodiode output voltage and the displacement of the piezoelectric positioner in the z direction $\Delta V/\Delta z_P$ is determined by completing an approach and retraction cycle and then measuring the slope of ΔV versus Δz_P when the tip is in contact with the substrate. This is usually completed at a relatively high contact force (~ 1500 pN). In these conditions the displacement of the cantilever Δz_C equals that of the piezoelectric positioner Δz_P . The slope $\Delta V/\Delta z_P$ is the optical lever sensitivity, C, and depends on the shape of the laser spot on the photodetector and the refractive index of the medium used.

The k_c allows the force applied to the cantilever to be calculated using Hooke's law. A number of models have been developed to estimate the spring constant of the cantilever [68]. The most common method employed for cantilever calibration for protein unfolding experiments is the thermal fluctuation method, which was first derived by Hutter *et al.* (1993). This model assumes that, in equilibrium, the cantilever tip behaves as a simple harmonic oscillator, with one degree of freedom fluctuating in response to thermal noise. This assumption is a good approximation when the deflections of a cantilever are small. The AFM tip is therefore modelled as an ideal spring with mass, m [69, 70]. When the tip is far from the sample, its motion will be dominated by thermal fluctuations [69] and the frequency of the motion near the resonant frequency will allow for an approximation of the k_c . The cantilever undergoes a deflection due to thermal motion in the vertical direction Δz_C , with an angular resonant frequency, ω_o . [69, 70].

$$\frac{1}{2}m\omega_0^2\langle z_C^2 \rangle = \frac{1}{2}k_c\langle z_C^2 \rangle \quad (2)$$

where

$$k_c\langle z_C^2 \rangle = k_B T \quad (3)$$

To obtain the k_c using this method, the cantilever is positioned within the buffer droplet at a distance from the surface. A power spectrum is taken (Figure 10) to measure the oscillation of the tip due to thermal fluctuations. This power spectrum is fast Fourier transformed and a Lorentzian¹ is fitted to the resonant peak. The area under the curve determines the mean square displacement of the cantilever due to thermal motion in units of volts.

For SMFS experiments the cantilever spring constant must be soft enough to respond to the piconewton forces observed for protein unfolding events, typically between 6 - 100 pNnm⁻¹. The force sensitivity of a cantilever is limited by its thermal noise [72] therefore cantilevers with low spring constants (softer) are more susceptible to noise[73]. Measurements using the AFM are limited by instrument stability and force sensitivity making the measurement of small forces over long periods of time challenging. Recent advances have been made to improve the precision of force measurements. For example, given the short-term force precision of a cantilever is sensitive to the hydrodynamic drag on the cantilever, by reducing the size of the cantilever, this drag can be reduced [74]. Reducing the length and thickness of the cantilever does not increase the spring constant but significantly reduces the noise of the cantilever. In another recent development, reducing the amount of reflective gold coating on the cantilever has led to significant improvements in the long-term force precision of a cantilever [75-77]. For example, removing some of the gold coating on a cantilever reduced the instrumental drift of soft cantilevers [75].

5.3 Polyproteins

In SMFS it is challenging to distinguish between protein unfolding events and nonspecific events, such as tip-surface interactions or sample contaminants. Therefore, a number of studies use polyproteins to provide a robust 'mechanical fingerprint' for protein unfolding. A polyprotein is a tandem array of single protein domains interlinked by polypeptide linkers (Figure 11A and B). Typically a polyprotein contains 5 - 8 protein domains[78]. These protein domains can be the same protein (homopolyprotein) or two or more different protein domains (heteropolyproteins or chimera polyproteins). The use of polyproteins domains

¹With functional form $P(f) = \frac{A}{(f-f_0)^2+B} + bg$, where f is the frequency, f_0 is the resonant frequency, bg corresponds to foreign noise with a uniform and constant spectrum, A and B are parameters that be used to determine the quality factor, Q ($Q = \frac{B}{2A^{1/2}}$) height of the peak and L_{peak} ($L_{peak} = \frac{A}{B}$) [69, 71])

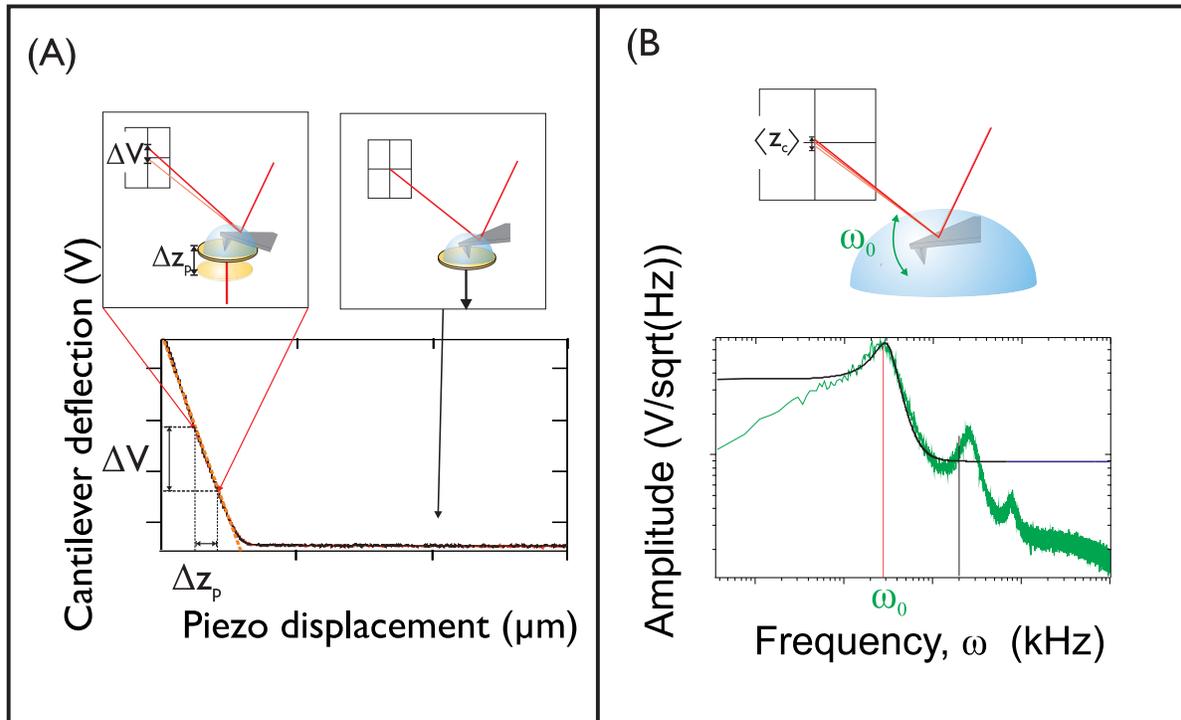


Figure 10: Cantilever calibration (A) The optical lever sensitivity is determined by completing an approach (red) and retraction (black) cycle between the cantilever tip to the substrate and then measuring the slope of cantilever deflection, as measured by the change in voltage ΔV of the photodetector, versus the piezoelectric positioner displacement, Δz_p , when the tip is in contact with the substrate. (B) The cantilever spring constant can be determined using the thermal noise method. The cantilever is positioned within the buffer droplet at a distance from the surface and a power spectrum is taken to measure the oscillation of the tip due to thermal fluctuations. The area under the curve in the power spectrum can be used to determine the mean square displacement of the cantilever, from which the spring constant can be calculated.

in SMFS has many advantages; it can increase the statistics of an experiment, decrease the interference of surface and nonspecific interactions and provide reference domains within the construct. Polyproteins provide a unique pattern in force extension traces. Over the last two decades a number of different techniques have been employed to engineer such polyproteins [8, 79–81]. A classical approach is based on the assembly of polymerase chain reaction (PCR)-generated DNA cassettes that together encode the full-length polyprotein [8, 82]. This method uses specific restriction sites between DNA fragments that encode respective protein domains. Sequential enzymatic digestion and ligation generate the full length polyprotein DNA in a stepwise manner. This method allows for the precise control of the number of domains in the polyprotein and the order of protein domains within a chimeric polyprotein. This method has been employed to make a wide range of different homo- and chimeric polyprotein constructs, making it a versatile method for polyprotein production. [8] However, in this approach the substitution of single DNA fragments occurs sequentially, making the process both laborious and time-intensive. An alternative method is based on the chemical coupling of identical protein monomers or dimers [81, 83, 83, 84]. Here, proteins can be linked through either disulfide bridge formation between cysteine pairs at designed locations or maleimide coupling of sulfhydryl groups within the protein. This chemical coupling allows for a faster one-step construction of polyproteins. This process also enables precise control of the pulling direction of proteins. However, this method precludes the generation of more complex protein scaffolds and generates an ensemble of low- to high-order multimers, which may have an impact on the subsequent analysis of the experimental data. Very recently we have developed a method which exploits Gibson Assembly (GA) (Figure 11C) and allows for rapid production of both homo- and heteropolyproteins of specific length and sequence, thereby retaining the benefits of the classical “cassette” approach described above (which results in a defined polyprotein composition) while addressing its previous drawback (expense in terms of labor and time). In GA the enzymatic assembly of DNA allows for the joining of many (up to 20) DNA fragments in a single step, using the combined function of three enzymes. The application of GA to generate polyprotein open reading frames compared to previous methods is advantageous because it provides the speed and ease of a single-step chemical coupling of protein monomers while allowing for control over type, order of domains, and length of the designed polyprotein.

5.4 Force-extension protein unfolding

In SMFS the AFM (Figure 9) is used to extend single protein molecules and measure the force responses of proteins from just a few up to hundreds of picoNewtons [85]. In a typical experiment, the polyprotein constructs are bound at one end to a substrate in solution. For example proteins are sometimes bound to a gold surface through cysteine residues which are genetically encoded onto one end of the polyprotein chain. At the other end, the polyprotein is picked up by a sharp tip, often silicon nitride, mounted on a flexible cantilever. Whilst it is unclear how the protein interacts with the silicon nitride tip it is often suggested that the polyproteins non-specifically adsorb to the tip via physisorption and application of a large contact force by the tip on the surface is hypothesised to increase the probability of adsorption, however, there is evidence to suggest that the interaction is electrostatic in nature [86]. Due to the nature of the interaction between the tip and the polyprotein, containing N domains, it can be picked up and extended anywhere along the chain, therefore between 0 and N unfolding events can be observed in a single force-extension cycle. Methods to functionalise the interaction between the tip and the protein have been employed to increase the likelihood of observing N unfolding events by precisely tethering the polyprotein between the two termini. For example, some groups have the engineered cysteine residues on one end of the polyprotein, and engineered an additional HaloTag protein on the other end. The HaloTag protein can covalently bind to chloroalkane ligands. These chloroalkane ligands can be covalently attached to a glass substrate leaving cysteine residues free to form a covalent bond with the cantilever tip [87, 88].

After the cantilever is calibrated, it is repeatedly brought into contact with the surface at a constant velocity (Figure 12). A large force (typically in the range 800 – 3000 pN) is applied to the surface in these experiments which is hypothesised to increase the probability of protein adsorption to the tip. In force-extension mode the tip is then retracted from the surface at a constant velocity. If a polyprotein was attached to the tip, a restoring force is measured. Initial resistance to extension in a polyprotein is a result of entropic forces, due to the entropic elasticity of unfolded parts of the polymer chain. This is a result of the preference of a chain to form a random coil to maximise its entropy. Stretching a polyprotein chain reduces its entropy, producing a restoring force that bends the cantilever. The entropic elasticity of a polymer follows

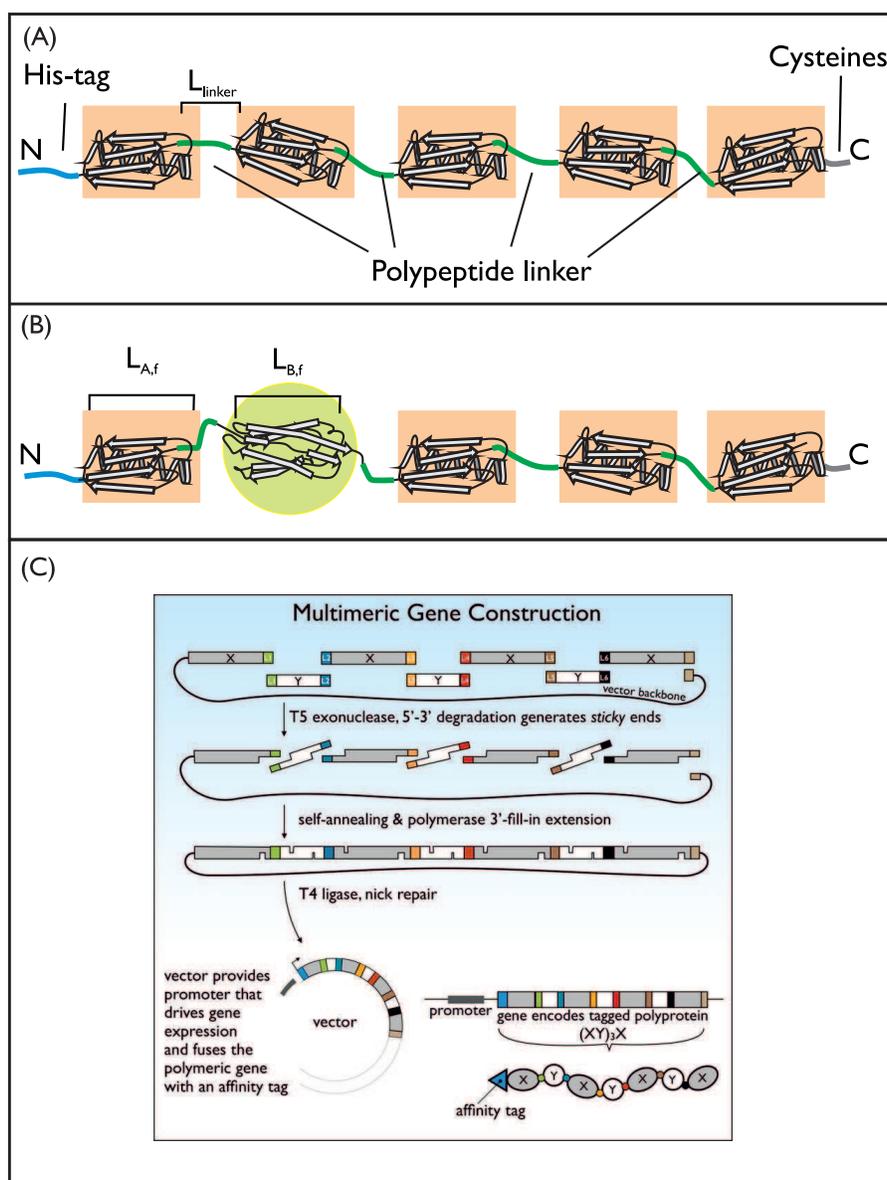


Figure 11: (A) Schematic of a homopolyprotein chain containing repeats of the same protein domain, polypeptide linkers of length L_{linker} , a His-tag (N terminus) and two terminal cysteines (C terminus). In a typical experiment in our research group, the polyprotein constructs are bound at one end to a gold substrate in solution, through the cysteine residues enabling the polyprotein to form a covalent bond with the gold substrate (an Au-S bond). (B) Polyproteins can be made which contain two protein domains. In this example protein B, with a folded length $L_{B,f}$ is sandwiched between protein A, with folded length $L_{A,f}$. (C) Gibson Assembly (GA) allows for rapid production of both homo- and heteropolyproteins of specific length and sequence. In GA the enzymatic assembly of DNA allows for the joining of many DNA fragments in a single step, using the combined function of three enzymes. It provides a single-step method for the coupling of protein monomers while allowing for control over type, order of domains, and length of the designed polyprotein. Printed with permission from [80]. Copyright 2015 American Chemical Society.

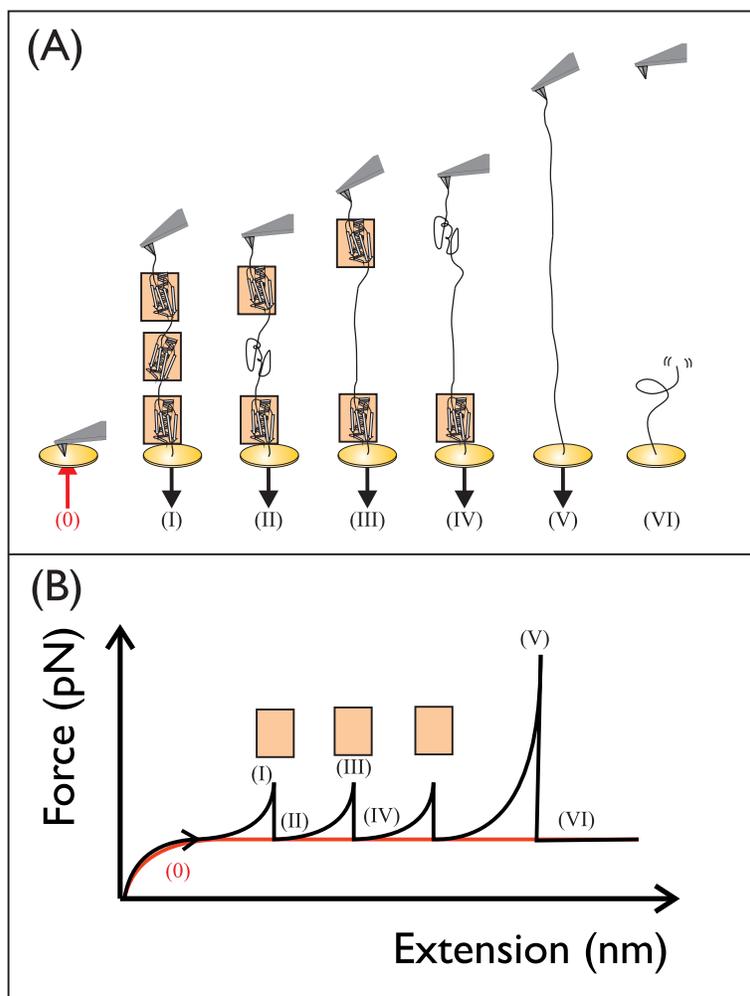


Figure 12: Schematic diagram showing the steps involved in obtaining a force-extension trace. (A) In force-extension mode the tip is retracted from the surface at a constant velocity (0). If a polyprotein is attached to the tip it is first elongated as the distance between the tip and surface increases (I). Further extension of the polyprotein causes it to unravel, typically in an all or nothing fashion (II). This unravelling releases amino acids, resulting in an increase in the end to end length of the polyprotein. At this point the forces acting on the cantilever abruptly decreases as the tension is lost and therefore the cantilever deflection changes. This unfolded protein domain will continue to be fully elongated adding to the effective length of the chain (III). A detachment peak is then observed which corresponds to the protein detaching from the tip or substrate (VI). (B) Schematic showing the associated features observed from (A). The red line is the approach of the substrate to the cantilever (0), when the cantilever is in contact with the surface it is deflected upwards causing the force to be negative initially. The retraction of the cantilever from the surface at constant velocity is shown as the black line. An increase in restoring force is observed as the distance between the tip and the surface increases. This force occurs as the protein resists the change in entropy of the chain. The bonds within the protein will be broken at a particular force (I). This will result in the elongation in the unfolded protein chain with no resistance followed by another restoring force on the cantilever as a force is applied across another folded protein domain (II). After all of the domains (three in this example) have unfolded the protein will detach either from the tip or substrate (V).

a nonlinear relationship which can be modelled by a number of polymer models (see later section). Further extension of the polyprotein causes it to unravel, typically in an all or nothing fashion. This unravelling releases previously condensed polypeptide chain, resulting in an increase in the end to end length of the polyprotein. At this point the forces acting on the protein drop as the tension is lost and the cantilever deflection changes. This unfolded protein domain will continue to be elongated adding to the effective length of the chain. Once fully extended, a force will be applied across the remaining folded protein domains and the process will begin again. This process is repeated until all the protein domains attached to the cantilever tip are unfolded. A detachment peak is then observed which corresponds to the protein detaching from the tip or substrate. This process is illustrated in Figure 12. The force-extension profile therefore appears as a saw-tooth pattern, where each peak reports on the unfolding of a protein domain. A single unfolding peak for each domain indicates a two-state unfolding process. These sawtooth patterns are characteristic of polyproteins and provide a mechanical fingerprint in SMFS experiments [82]. Furthermore, the mechanical fingerprints of new proteins can be obtained by sandwiching uncharacterised proteins between previously characterised proteins [89–91].

5.5 Analysis of force-extension data

Force extension curves from single molecule force experiments are typically filtered, for example to only include traces where 2 to N unfolding events are observed, where N is the total number of protein domains in the homopolyprotein (made up of only 1 protein type), assuming 2-state behaviour. The traces are also filtered to only include those with minimal surface interactions close to the surface and a clear detachment peak [92]. Figure 12 shows how a force-extension trace is obtained. The red line is the force measured during the approach of the cantilever to the surface. The black line is the force measured upon the retraction of the cantilever from the surface at constant velocity. In the resulting graph, each peak is an unfolding peak and the last peak is the detachment of the elongated chain from the cantilever and is known as the detachment peak. The protein chain can also dissociate from the cantilever before detachment occurs, this typically results in a smaller detachment peak. Experiments are completed a number of times at a range of retraction velocities, to allow a pulling speed dependence on the mechanical stability to be obtained. By collecting data sets at multiple velocities within a single experiment, errors due to cantilever calibration can be minimised. The forces from this single experiment can be compared with forces collected for the same experimental velocity but obtained on different days, if forces are not consistent with those previously observed then the cantilever may be missed calibrated.

The force-extension curves of a homopolyprotein reveal a series of equally spaced force peaks as the distance between the tip and the substrate increases. In the case of a simple all or nothing, two-state unfolding protein, each force peak corresponds to the unfolding of a single protein domain. A consistent stepwise increase of the chain length is observed between two consecutive unfolding events. The interpeak distance x_p , corresponds to the elongation of a protein domain upon unfolding (Figure 13). x_p gives a measure of the mechanical structure of the domain, since it is related to the number of amino acids initially ‘trapped’ between the bonds that form the main mechanical resistance to unfolding and subsequently ‘released’ upon unfolding. The unfolding force, F_{UN} , of a protein is the peak force reached before the protein unfolds. This gives a measure of the mechanical stability of a protein. Forced unfolding of a protein in SMFS occurs as a result of the balance between thermal and mechanical forces. Unfolding is therefore a stochastic process and a distribution of F_U values for any unfolding event is expected. By obtaining many force-extension curves, measuring a number of F_U and creating a histogram of unfolding forces, the most probable unfolding force can be obtained at a particular pulling velocity. This can be visualised in a scatter plot of the x_p and F_U from a number of force-extension curves at the same pulling speed (Figure 13). Whilst, often the F_U is taken to be the maximum force before the protein unfolds, more sophisticated methods have been developed to determine the exact extension that mechanical rupture occurs. One method converts extensions, which is the tip-sample separation, into the total length of the protein, the contour length. The extension in an experiment is sensitive to the loading rate and other experimental factors. The extension can be converted to contour length by solving polymer models describing forced extension such as the worm-like chain model (see Section 6.2) for contour length at a given force and extension and accounting for backbone elasticity of the polypeptide chain [93, 94]. Superposition of force-contour length graphs are used to form histograms of barrier position (point of rupture); this method allows direct determination of the barrier positions within

the protein and the corresponding F_U . More recently this analysis has been automated to align the force-contour length traces and group them according to number of unfolding events observed [95]. Away from the unfolding events, the other sections of the force versus extension profiles can be further understood using polymer elasticity models.

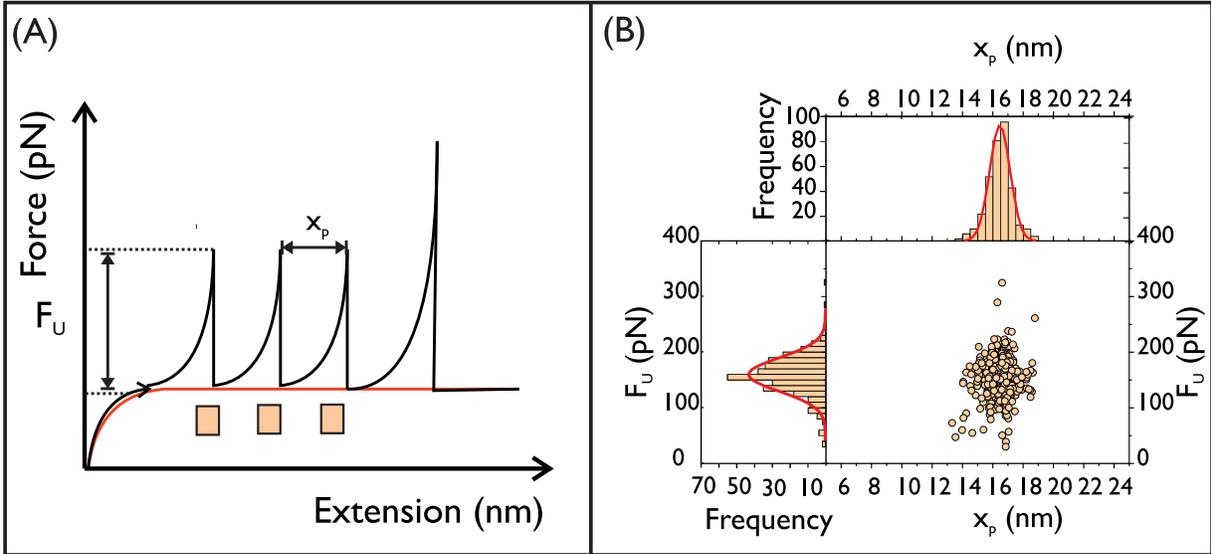


Figure 13: (A) The force-extension curve of a polyprotein is a series of equally spaced force peaks as the distance between the tip and the substrate increases. The interpeak distance, x_p , corresponds to the elongation of a protein domain upon unfolding. The unfolding force, F_U , of a protein is the peak force reached before the protein unfolds. This gives a measure of the mechanical stability of a protein. (B) By measuring the x_p and F_U from a number of force-extension curves at the same pulling speed, a scatter plot of unfolding forces can be obtained. This allows for the identification of a ‘mechanical fingerprint’ of the protein of interest.

6 Stretching proteins: polymer elasticity models

6.1 The freely jointed chain model

Stretching a polyprotein chain reduces its entropy, producing a restoring force that results in bending of the cantilever. The entropic elasticity of a protein follows a nonlinear relationship that can be formally described by a number of models of polymer elasticity. One simple model for polymer elasticity is known as the freely jointed chain (FJC) model or ideal chain model [96, 97]. This model was initially proposed by Kuhn in 1936 [98]. In this model, a polymer is partitioned into N segments each with length, b . This length, b , characterises the Kuhn length of the polymer. The maximum total length of this polymer, or contour length, L_c is therefore calculated as

$$L_c = \sum_{i=1}^N b = Nb \quad (4)$$

The projection of each of the segments is independent of the preceding and proceeding segments. This model therefore does not account for the long range interactions of a chain, and the chain can undergo self-interactions. Each segment, i , has an associated vector, r_i , the mean square end to end distance, $\langle R^2 \rangle$, is then:

$$\langle R^2 \rangle = \langle \mathbf{R} \cdot \mathbf{R} \rangle = \left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \sum_{i=1}^N \mathbf{r}_i \right\rangle \quad (5)$$

$$= \sum_{j=1}^N \sum_{i=1}^N \langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle = \sum_{i=1}^N \langle \mathbf{r}_i \cdot \mathbf{r}_i \rangle + \sum_{i=1}^N \sum_{j \neq i}^N \langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle = Nb^2 \quad (6)$$

Where $\sum_{i=1}^N \sum_{j \neq i}^N \langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle$ is a consequence of the segments being uncorrelated. If the chain is fixed at one end and subjected to a constant force, F , at the other end, an entropic restoring force will be experienced due to a reduction in the number of available states to the chain (Figure 14). The energy of this FJC (E_{FJC}) can be determined by the work done ($-W$) to change the shape of the polymer.

$$E_{\text{FJC}} = -W = -\mathbf{F} \cdot \mathbf{R} = -F \left(\sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{z} \right) = -Fb \sum_{i=1}^N \cos \theta_i \quad (7)$$

Where each segment has an associated angle, θ_i with the direction of the applied force. The partition function, Z , is computed from summing the energies from all configurations (conf.) of the protein:

$$Z = \sum_{\mathbf{r}_i, \text{conf.}} \exp(-E_{\text{FJC}}/k_B T) = \sum_{\mathbf{r}_i} \prod_{i=1}^N \exp(Fb \cos \theta_i / k_B T) \quad (8)$$

As each segment is independent and therefore uncorrelated with the others, the energies will be the same.

$$Z = \left[\int_0^{2\pi} d\psi \int_0^\pi \sin \theta d\theta \exp\left(\frac{Fb \cos \theta}{k_B T}\right) \right]^N = \left[\frac{4\pi k_B T}{Fb} \sinh(\theta) \right] \quad (9)$$

The free energy, H , of this system can be computed as $H = k_B T \ln Z$. Therefore the average displacement, $\langle x \rangle$ can be computed from:

$$\langle x \rangle = \frac{d(k_B T \ln Z)}{dF} = \frac{d}{dF} \left(N \ln \left(\frac{4\pi k_B T}{Fb} \right) + N \ln \left(\sinh \frac{Fb}{k_B T} \right) \right) = L_c \left(\coth \frac{Fb}{k_B T} - \frac{k_B T}{Fb} \right) \quad (10)$$

At small forces, the polymer behaves as a Hookean spring with $F = \frac{3k_B T}{b} \frac{x}{L}$. Whilst this model can describe the reaction of biological molecules to forces at small extensions, it fails to describe the elastic response at high forces for the stretching of DNA [99].

6.2 The worm-like chain model

The worm-like chain model (WLC) can be used to describe the semi-flexible behaviour of polymers. It is an extension to the FJC model of polymer elasticity in the continuous limit as $b \rightarrow 0$. The model predicts the entropic restoring force, F , of the polyprotein at any given extension from two parameters: L_c the contour length (the maximum total length of the unfolded polyprotein) and p the persistence length (the length of the smallest rigid component of the chain). In the WLC model (Figure 15, the polymer is described by a continuous chain of N segments, each with length d , that are at an angle θ to the adjacent segment. The projection of each segment, from the first segment to the last, affects the projection of the adjacent segment along the chain by $l \times \cos(\theta)$. The projection onto the first bond is given by $\langle \cos(\theta_{1N}) \rangle = \cos^{N-1}(\theta)$. The average projection is therefore given by [97]:

$$\langle x \rangle = \frac{1}{d} \sum_{i=1}^N \langle \mathbf{d}_i \cdot \mathbf{d}_j \rangle = d \sum_{i=0}^{N-1} \cos^i(\theta) = \frac{d(1 - \cos^N(\theta))}{1 - \cos(\theta)} \quad (11)$$

As N tends to infinity this expression tends to [97]:

$$\lim_{N \rightarrow \infty} \langle x \rangle = \frac{d}{1 - \cos(\theta)} = p \quad (12)$$

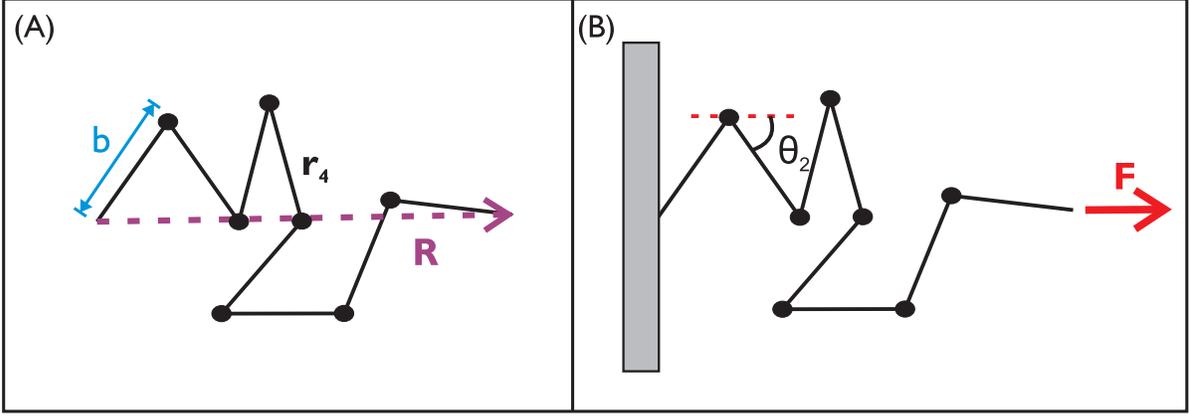


Figure 14: (A) The most simple model for polymer elasticity is known as the freely jointed chain (FJC) model or ideal chain model. In this model, a polymer is partitioned into N segments each with length b . This length b characterises the Kuhn length of the polymer. The distance \mathbf{R} represents the end to end vector length of the chain. (B) If the chain is fixed at one end and is subjected to a constant force, \mathbf{F} , at the other, an entropic restoring force will be experienced due to a reduction in the number of available states to the chain. Each chain makes an angle θ_i to the direction of the applied force in this diagram θ_2 is shown which represents the angle for the second segment in the chain. The energy of this FJC can be determined by the work done to change the shape of the polymer.

This is the definition of p , a measure of polymer stiffness which determines the distance over which the correlations between the segment are lost. For a flexible and “bendy” chain, the correlations are lost quickly, and therefore p is small. In a more rigid rod-like chain the p will be larger because the chain will be correlated for larger distances along the chain. For example WLC fits to the force-extension traces of a β -sheet protein gave a persistence length of about 0.4 nm [79]. This roughly corresponds to the length of a single amino acid. However, in more recent studies it has been suggested that this measured value of p accounts for both entropic and attractive interactions due to hydrophobic contacts within a protein [100].

For small angles $\cos(\theta) \approx 1 - \frac{\theta^2}{2}$ the expression given in Equation 12 tends to $p = \frac{2d}{\theta^2}$. In the WLC model the segment length becomes infinitely small as the number of segments in the chain increases, whilst the contour length, $L_c = Nd$, and the p remain constant. By replacing $\cos^N(\theta) \approx (1 - \frac{\theta^2}{2})^N$ by $\exp(-N\theta^2/2)$ the projection in this limit then becomes [97]:

$$\langle x \rangle = p(1 - e^{-\frac{x}{p}}) \quad (13)$$

The exponential term means that the angular correlation decays exponentially along the chain. As $L \rightarrow \infty$ this would tend to the persistence length of the chain. The mean square radius of gyration is also given by:

$$\langle r^2 \rangle = 2p \left[1 - \frac{p}{L} (1 - e^{-\frac{L}{p}}) \right] \quad (14)$$

There is an energy cost associated with the bending of a WLC given by a bending elasticity term [101]:

$$E = \frac{\kappa}{2} \int_0^L \left(\frac{\partial^2 \hat{\mathbf{r}}(s)}{\partial s^2} \right)^2 ds \quad (15)$$

This term accounts for the shape of the polymer. Where E is the bending energy, s is the distance along the curve, κ is the bending coefficient and $\frac{\partial^2 \hat{\mathbf{r}}(s)}{\partial s^2}$ is the tangent vector along the chain [101–103]. To straighten out the chain, work needs to be done against the bending energy (Figure 15). The Hamiltonian, E_{WLC} , for such a system is given by Equation 16 [102], where $\mathbf{F} \cdot \mathbf{R}$ accounts for the application of a constant force to the ends of the chain.

$$E_{\text{WLC}} = \int_0^L ds \frac{\kappa}{2} \left[\frac{\partial^2 \hat{\mathbf{r}}(s)}{\partial s^2} \right]^2 - \mathbf{F} \cdot \mathbf{R} \quad (16)$$

Obtaining a force-extension relationship from this equation requires computing the force over all possible shapes of a polymer. The calculation of the extension can be computed similarly to the calculation for the ground state of a quantum mechanical dipole in an electric field [104]. This was first noted by Marko and Siggia in 1995 and an outline to the path integral procedure has been previously presented in a review [17]. The interpolation formula for this model has successfully been used to describe the unfolding of DNA by Marko and Siggia and Bustamante *et al.* and is given in Equation 17 [104, 105].

$$F(x) = \frac{k_B T}{p} \left[\frac{1}{4} \left(1 - \frac{x}{L_c} \right)^{-2} - \frac{1}{4} + \frac{x}{L_c} \right] \quad (17)$$

The force (F) is measured as a function of extension (x). The WLC can also be used to calculate the ΔL_c between protein domains in a polyprotein and this is typically is the change in length between the folded and unfolded protein. The WLC model has been used extensively to fit force-extension curves from forced protein unfolding [78, 79, 91, 106, 107]. Despite being able to describe the force-extension behaviour of most biomolecules, such as the stretching of DNA [105], this model fails for large forces (Figure 16) where over-stretching of the bonds within the polymer occurs [102]. Furthermore this model assumes that the persistence length is independent of extension and is identical for the folded and unfolded protein. However, it has been suggested that this is a crude estimation because the stiffness of a completely unfolding protein domain would differ from that of a fully folded protein [108]. Some alternative models have attempted to modify the WLC model to minimise the discrepancies between the model and the experimental data.

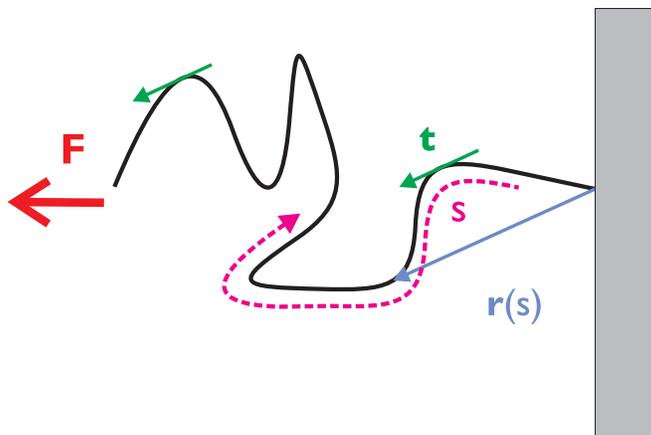


Figure 15: Schematic of a worm-like chain s where $\mathbf{r}(s)$ is the vector whose origin can be taken from any point in space and describes the shape of the chain. The vector \mathbf{t} is the unit tangent vector to the curve at any point. The chain opposes the change in shape as an applied force \mathbf{F} is applied to the ends.

6.3 Further developments in polymer elasticity models

Given the FJC and WLC models fail to describe the over-stretching of molecules at high force [102] there has been considerable effort to develop new models of force versus extension behaviour. Here we highlight some of these models which have particular relevance to protein unfolding. Beginning with developments of the FJC model, Smith *et al* have developed the extended FJC model which includes the elasticity of

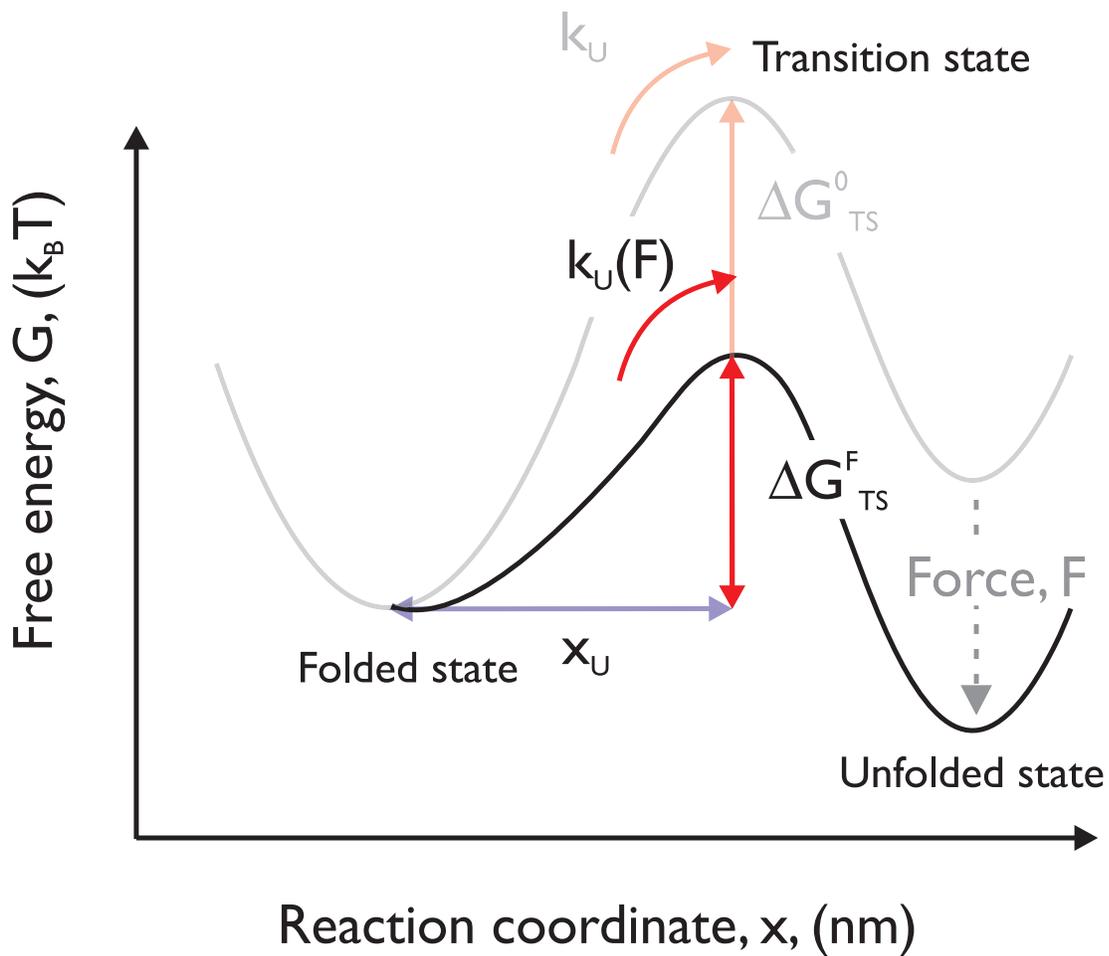


Figure 16: Examples of the use of polymer elasticity models to fit to protein unfolding data from SMFS force versus extension experiments. (A) the freely jointed chain (FJC) model is shown, with a Kuhn length of 0.22 nm for the force extension curve of the polyprotein (pL)₅. (B) The worm like chain (WLC) model is shown, with a persistence length, p , of 0.39 nm for the same data, (pL)₅. (C) A comparison of the model fit to the data for one protein unfolding event in the polyprotein chain for the FJC (blue) and WLC (red) models with an p of 0.4 nm and $b \approx 2p = 0.8nm$.

Model	Associated formula	Refs
Extended FJC model	$x = L_c \left(\coth \frac{Fb}{k_B T} - \frac{k_B T}{Fb} \right) \left(1 + \frac{F}{S} \right)$	[109]
Heterogeneous FJC model	$x = N_f L_{fs} \left(\coth \frac{Fb_f}{k_B T} - \frac{k_B T}{Fb_f} \right) + N_u L_{us} \left(\coth \frac{Fb_u}{k_B T} - \frac{k_B T}{Fb_u} \right)$	[108]
Extended WLC model	$F(x) = \frac{k_B T}{p} \left[\frac{1}{4} \left(1 - \frac{x}{L_c} + \frac{F}{\kappa_s} \right)^{-2} - \frac{1}{4} + \frac{x}{L_c} - \frac{F}{\kappa_s} \right]$	[110]
Exact solution to WLC model	$F(x) = \frac{k_B T}{p} \left[\frac{1}{4} \left(1 - \frac{x}{L_c} \right)^{-2} - \frac{1}{4} + \frac{x}{L_c} + \sum_{i=2}^{i \leq 7} \alpha_i \left(\frac{x}{L_c} \right)^i \right]$	[102]
Thick chain model	$F = \frac{k_B T}{a(1-x)} \tanh \left(\frac{k_1 x^{3/2} + k_2 x^2 + k_3 x^3}{1-x} \right)$ $k_1^{-1} = -0.28314 + \frac{0.76441\Delta}{a} + \frac{0.31858\Delta^2}{a^2}$ $k_2^{-1} = 0.15989 - \frac{0.50503\Delta}{a} - \frac{0.20636\Delta^2}{a^2}$ $k_3^{-1} = -0.34984 + \frac{1.23330\Delta}{a} + \frac{0.58697\Delta^2}{a^2}$	[111]
Freely rotating chain model	$\frac{x}{L_c} = 1 - \left(F_{WLC}(Fb/k_B T)^{-1} + \left(\frac{cFb}{k_B T} \right)^\beta \right)^{-1/\beta} + \frac{f}{\lambda}$	[112]

the segments in the polymer chain [109]. Su *et al.* have developed the heterogeneous FJC model which incorporates the difference in stiffness between the folded and unfolded protein chain [108]. In this model, the FJC model is extended to be the sum of the contribution of the number of folded and unfolded domains in a polyprotein. The chain is modelled as a heterogeneous FJC with two possible values of Kuhn length and contour length representing its folded and unfolded configurations. The authors obtained analytical solutions for the force–extension response of the polyprotein for different types of loading conditions and fit the analytical solutions to constant-velocity data for the protein ubiquitin. Extensions have also been made to the classical formula for the WLC model. The work by Wang *et al.* (1997) incorporates a stretching term in the WLC model [110, 113] to create the WLC model with segment elasticity (or extendable WLC model). In this model the ratio of extension to contour length, x/L_c , is replaced by $x/L_c + F/\kappa_s$, where κ_s is the extension modulus of the protein. Bouchiat *et al.* (1999) have incorporated additional terms to the WLC model [102]. By determining an exact solution to the WLC model and adding correction terms, they were able to eliminate discrepancies. This method was successfully implemented to reduce the errors between the experimental data for DNA stretching and the model. Models have also attempted to incorporate information about the 3D structure of a protein. One example is the thick chain model [111] which is used to infer information about the width of the chain in force-extension relationships. In a study by Thirumalai *et al.* a unified theory was developed to account for the high force entropic elasticity of biopolymers solely in terms of the persistence length and the monomer spacing [114]. This allows biopolymers to be treated as FJCs in one force regime and WLCs in another force regime. This approach is attractive as, by probing the response of biopolymers over a wide range of forces, the force dependent elasticity can be fully described.

7 Probing the effect of force on the kinetic and thermodynamic protein unfolding process

7.1 Dependence of the unfolding force on the loading rate

The force at which an protein unfolds depends on the loading rate used in the SMFS experiment.

$$r = \frac{dF}{dt} = k_c v \quad (18)$$

where k_c is the spring constant and v is the pulling speed. Given that forced unfolding of a protein occurs under the influence of two forces, mechanical and thermal forces, different pulling speeds reflect different time windows during which thermal forces can act under application of mechanical force. At high pulling speeds thermal forces have less time to act, resulting in a higher measured F_U (Figure 18). By performing force-extension SMFS experiments on the same construct at different velocities, the pulling speed dependence

can be obtained (Figure 18). The range of pulling speeds used in SMFS experiments vary in the range of a few nm/s to $\sim 10 \mu\text{m/s}$. The lower limit is due to the large mechanical loop of the AFM apparatus, which makes thermal drift the limiting factor. At high pulling speeds viscous drag becomes a problem [115–118]. At speeds above $10 \mu\text{m/s}$, the speed-dependent hydrodynamic drag force acting on a cantilever is of the same magnitude as the unfolding force measured on single biological molecules [116]. However the use of short, low-noise cantilevers and optimised design of the piezoelectric positioner can reduce this effect. Recently a high speed AFM has been developed to unfold proteins at velocities of $\sim 4 \text{ mm/s}$ [119]. This was achieved by using short cantilevers with small viscous damping ($0.035 \text{ pN } \mu\text{m}^{-1} \text{ s}^{-1}$) and the use of a miniature piezoelectric actuator. This development is significant because it now allows SMFS experiments to be completed at velocities which are equivalent to those probed in molecular dynamics simulations which offer atomic-level descriptions of the forced unfolding. In the pulling speed dependence plot in Figure 18 the schematic shows a linear dependence of the F_U with the natural logarithm of the velocity. Can we extrapolate this behaviour to higher and lower pulling velocities? Experimental data, including those from the recent fast-AFM study mentioned above, and theoretical studies suggest this may not be the case[119–121]. For example, in the high pulling speeds achieved using the fast-AFM a deviation from linear behaviour was observed at high pulling speeds [119]. Earlier SMFS experiments on the protein filamin, probed unfolding at low speeds found a deviation from linear behaviour in the F_U versus speed dependence[120]. The study proposed that the deviation could only be explained by a switch between parallel unfolding pathways and suggested that mechanical unfolding kinetics needs to account for the multidimensionality of the free-energy landscapes of proteins, which are crucial for understanding the behavior of proteins under the small forces experienced in vivo. In another study, the choice of the one-dimensional reaction coordinate probed in SMFS has been questioned due to the multidimensionality of protein energy landscapes[122]. It has been suggested that more information could be obtained about the multidimensional energy landscape from analysing the dynamical fluctuations of proteins using simulations and single molecule techniques [122].

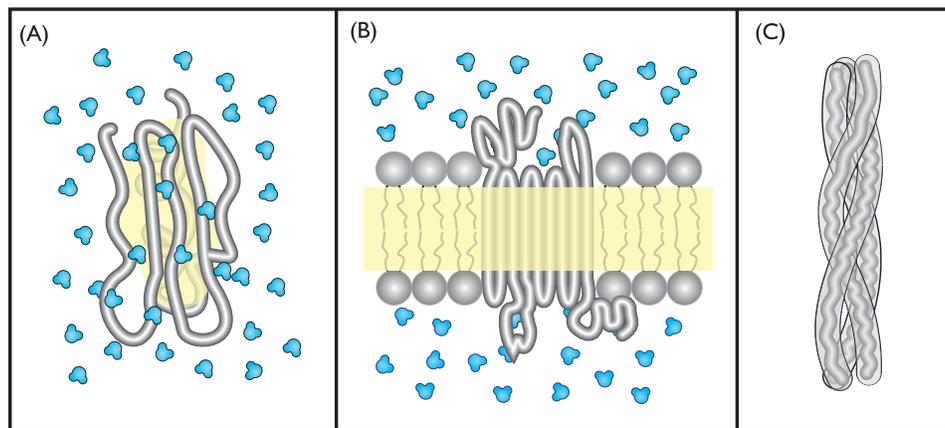


Figure 17: Recent development of a high speed AFM for SMFS force spectroscopy (A) The high speed AFM setup. An objective focuses the beam of the super luminescent light-emitting diode (SLED) and collects the light reflected by the cantilever, finally detected by a segmented photodiode. Titin I27 (concatemers of eight domains) are immobilized on a tilted gold-coated surface via C-terminal cysteines. They are pulled by their N-terminal histidine tag with a nickel-coated tip at the end of a short cantilever. Tilting the sample surface further reduces hydrodynamic forces. Top inset shows a titin I27 (also referred to as I91) domain (PDB ID: 1TIT) with relevant β strands coloured in blue (strand A), yellow (strand A'), green (strand B), and red (strand G). Bottom inset shows a scanning electron micrograph of a short cantilever. (B) Force-extension curves acquired at three different retraction velocities: 1, 100, and $1000 \mu\text{m/s}$. The $1 \mu\text{m/s}$ curve is moving average-filtered (red trace, $65\text{-}\mu\text{s}$ time window). Times to unfold a single I27 domain are indicated by arrows. From [119]. Reprinted with permission from AAAS.

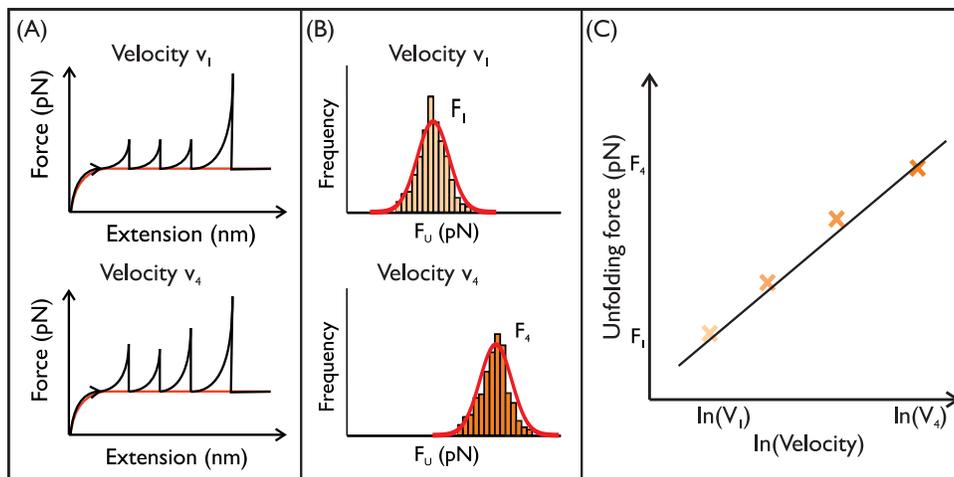


Figure 18: (A) SMFS force extension experiments can be completed at different pulling speeds. (B) By measuring the unfolding force of each force peak and creating an unfolding force histogram, the unfolding force F_U can be obtained for a particular pulling velocity. The higher the pulling velocity the higher the value of F_U . (C) Completing force versus extension experiments at a range of pulling velocities allow the F_U to be plotted as a function of the natural logarithm of the pulling velocity.

7.2 Kinetic parameters of protein mechanical unfolding

A number of models have been developed to relate the force response of proteins to the unfolding rate and height of the barrier separating the folded and unfolded states of a protein [120, 123–125]. The Bell-Evans-Richie model has been most extensively used to determine the free energy landscape of proteins using SMFS protein unfolding data [78]. Other models have emerged which incorporate more detail, and can account for the curvature that has been observed in the plot of force against the natural logarithm of the velocity [67, 126]. We will present two of these models: the Friddle-De-Yoreo and Dudko-Hummer-Szabo models. A comparison of the Bell-Evans-Richie, Friddle-De-Yoreo and Dudko-Hummer-Szabo models on protein binding forces has recently been performed [127]. In this study SMFS experiments measured the binding forces of amyloid- β (1-41) peptides and compared three different theoretical models used to fit the experimental data. Interestingly the results showed that the Friddle-De-Yoreo and Dudko-Hummer-Szabo models yielded similar ΔG_{TS}^0 to one another but differed significantly from the results obtained for the Bell-Evans-Richie model.

7.2.1 The Bell-Evans-Richie Model

Protein unfolding via a two-state process can be described by an energy barrier between the folded and unfolded states of the protein. This 1D energy barrier is defined by the reaction coordinate studied, which in an SMFS experiment is the end-to-end length of the protein. The Bell-Evans-Richie model is an empirical model, which can be applied to data obtained from SMFS that describe the rate of unfolding of a biological molecule under a mechanical force. The model originates from George Bell [123] who described the rate of adhesion of cells to other cells or surfaces in the presence of hydrodynamic stress [123]. The kinetic model for mechanical strength was first derived by Zhurkov who applied the same principles to solids under mechanical strength [128]. The Bell-Evans-Richie model suggests that the rate of unfolding, under applied force, decreases exponentially with force. Evans and Richie [129] provided the theoretical framework to aid in the understanding of this model for protein unfolding by deriving Equation 20 from Kramer’s reaction rate theory for a protein in a deep harmonic well with high barriers and friction to unfolding [130]. Kramer’s theory describes the mechanical unfolding as diffusion by Brownian motion over a 1D energy profile probed as a function of extension of the protein. In the derivation provided by Evans and Richie, the prefactor contains

information about the curvature of the energy well of the native and transition states and the friction along the 1D reaction coordinate [131]. However, these parameters are difficult to determine experimentally and therefore, it is typically assumed to be unperturbed by force.

$$k_U(F) = A \exp \frac{-(\Delta G_{TS}^0 - Fx_U)}{(k_B T)} \quad (19)$$

$$= k_U \exp \frac{(Fx_U)}{(k_B T)} \quad (20)$$

Where k_U is the rate of unfolding, ΔG_{TS}^0 is the height of the unperturbed free energy barrier opposing the mechanical fracture (or activation energy to the transition state) and x_U is the distance that separates the native state of the protein and the transition barrier. The attempt frequency, A , is typically assumed to be an arbitrary value which ranges from a fixed value of $10^6 - 10^9$. Using the Bell-Evans-Richie model, the k_U and x_U can be directly determined from the associated $k_U(F)$, the rate of unfolding under applied force. The rate of unfolding is force dependent and the transition barrier height is perturbed by an amount Fx_U . By applying a force across the protein, the energy barrier to the unfolded state decreases (Figure 19). This increases the probability that the protein will unfold at a certain force. Using the Bell-Evans-Richie model, the rate of unfolding can be related to the force applied. If it is assumed that the transition state and native state on the 1D free energy profile are independent of force, then the activation barrier at any force is given by $\Delta G_{TS}(F) = \Delta G_{TS}^0 - F \times x_U$. Force therefore lowers the energy barrier by tilting the energy landscape, (see Figure 19) assuming there is no change in the shape of the energy landscape [126]. The largest contribution to the change in the kinetics under applied force is due to a change in the probability of barrier crossing, therefore the change in shape and location of the transition barrier has been shown to be insignificant provided the energy barriers are sharp [129]. In constant velocity experiments, the force applied to the protein constantly changes as a function of time. This defines the loading rate of the protein which probes the lifetime of the non-covalent interactions within proteins as the force is varied. The distribution of unfolding forces, F_U can be shown to be described by the probability distribution[131]:

$$P(F_U) = \frac{k_U(0)}{R} \exp \frac{F_U x_U}{k_B T} \exp \left(\frac{k_U(0) k_B T}{R x_U} \left[1 - \exp \frac{F_U x_U}{k_B T} \right] \right) \quad (21)$$

This formula can be used to fit the distribution of unfolding forces observed in SMFS experiments. The most probable force of unfolding can also be determined from the maximum of $P(F)$ and is given by Equation 22[131]. This describes how the most probable force, F_U^* , is affected by the loading rate.

$$F_U^* = \frac{k_B T}{x_U} \ln \frac{R x_U}{k_B T \times k_U} \quad (22)$$

7.2.2 Monte Carlo simulations and the Bell-Evans-Richie model

Using the analytical solution in equation 22 it should, in principle, be possible to calculate the kinetic parameters of protein unfolding, k_U and x_U . However, in the experiment the loading rate, Equation 18, varies as the molecule is stretched because the stiffness of the molecule depends on the applied force. Therefore Equation 22 cannot be used directly. Instead, k_U and x_U can be determined using Monte Carlo (MC) simulations that mimic the stochastic nature of the thermally driven unfolding of the protein. In the MC simulation, details of the experiment are input, including the average persistence length, average spring constant, the linker length, the length of a single folded and unfolded protein and the temperature of the experiment [90, 132]. The unfolding rate under an applied force, k_U , is calculated using the Bell-Evans-Richie model and then the probability of a single domain unfolding, P_U , at a defined time step of Δt is

$$P_U = N k_U(F) \Delta t \quad (23)$$

where N is the number of folded domains in the polyprotein and $k_U(F)$ is the affect of force on the unfolding rate. The simulation is performed for the full range of experimental pulling speeds (Figure 18C) using different combinations of k_U and x_U . The simulations are used to generate force distributions for each experimental pulling velocity, which can be compared to the experimental force distribution.

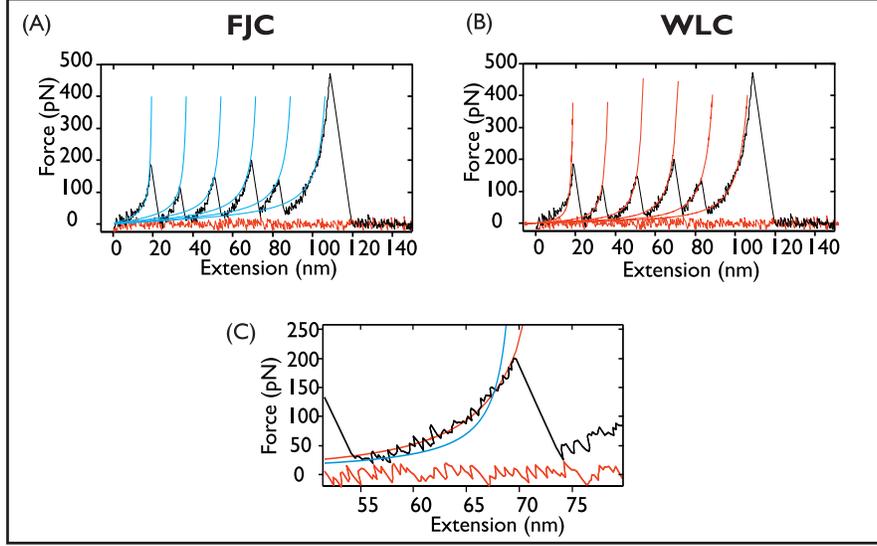


Figure 19: Schematic showing the effect of mechanical force on the energy landscape of a protein. The free-energy profile of a protein folding pathway leading from the native, folded state to its denatured, unfolded state through a transition state at an unfolding rate k_U ; a simple two-state transition. The application of an external force tilts the energy landscape and thus lowers the energy barrier to arrive at the unfolded state, ΔG_{TS}^F . Accordingly, a sufficiently high force deforms the profile such that eventually the unfolded state becomes more populated than the folded state.

7.2.3 Friddle-de-Yoreo model

A recent study challenged the assumption made by the Bell-Evans-Richie model that no reversible refolding occurs during protein unfolding experiments[125]. In the Friddle-de-Yoreo model, a system of two bodies being pulled apart passes through two phases, an equilibrium phase at lower pulling velocities where the molecules can rebind, and a kinetic phase at higher loading rates where molecule unbind irreversibly. The equilibrium phase is associated with a shallow slope and the kinetic phase with a steeper slope leading to the nonlinearity observed in force- $\ln(\text{rate})$ graphs. The model therefore accounts for the probability that reversible binding or protein refolding can occur, particularly in the limit of lower rates. This model allows both kinetic and thermodynamic information about the interactions occurring in the rupture event to be accounted for. The force F_{eq} , at which the phase changes from equilibrium to kinetic is given by:

$$F_{eq} = \sqrt{2k_c \Delta G_U^0} \quad (24)$$

F_{eq} is dependent of the cantilever spring constant, k_c and the free energy of unfolding ΔG_U^0 . In this model both the rate of unfolding, k_U and the rate of refolding k_F are taken into account.

$$k_U(F) = k_U \exp \left[\beta \left(F x_U - \frac{1}{2} k_c x_U^2 \right) \right] \quad (25)$$

$$k_F(F) = k_F \exp \left[-\beta \frac{k_c}{2} \left(\frac{F}{k_c} - x_U \right) \right] \quad (26)$$

Where F is the unfolding force, x_U is the distance between the unfolded state and the barrier to unfolding, k_U is the rate of unfolding at zero force and k_F is the rate of refolding at zero force. $\exp(\beta \Delta G_U^0) = k_F/k_U$ is used to define the activation energy, ΔG_U^0 . An approximation can be made for the unfolding force, $F_U(r)$ by:

$$\langle F_U(r) \rangle = F_{eq} + \frac{k_B T}{x_U} \ln \left(1 + \frac{R e^{-\gamma}}{k_U(F_{eq}) \frac{k_B T}{x_u}} \right) \quad (27)$$

Where γ is Euler's constant. This can be used to fit the force versus loading rate relationships. However, it is challenging to accurately determine the loading rate for polyproteins. The loading rate depends on the effective spring constant, which is a combination of the spring constant of the cantilever and protein. As a protein unfolds the effective spring of the chain changes, as it is the summation of folded and unfolded domains. Due to the stochastic nature of the number of unfolding events, the loading rate is not constant between force-extension traces.

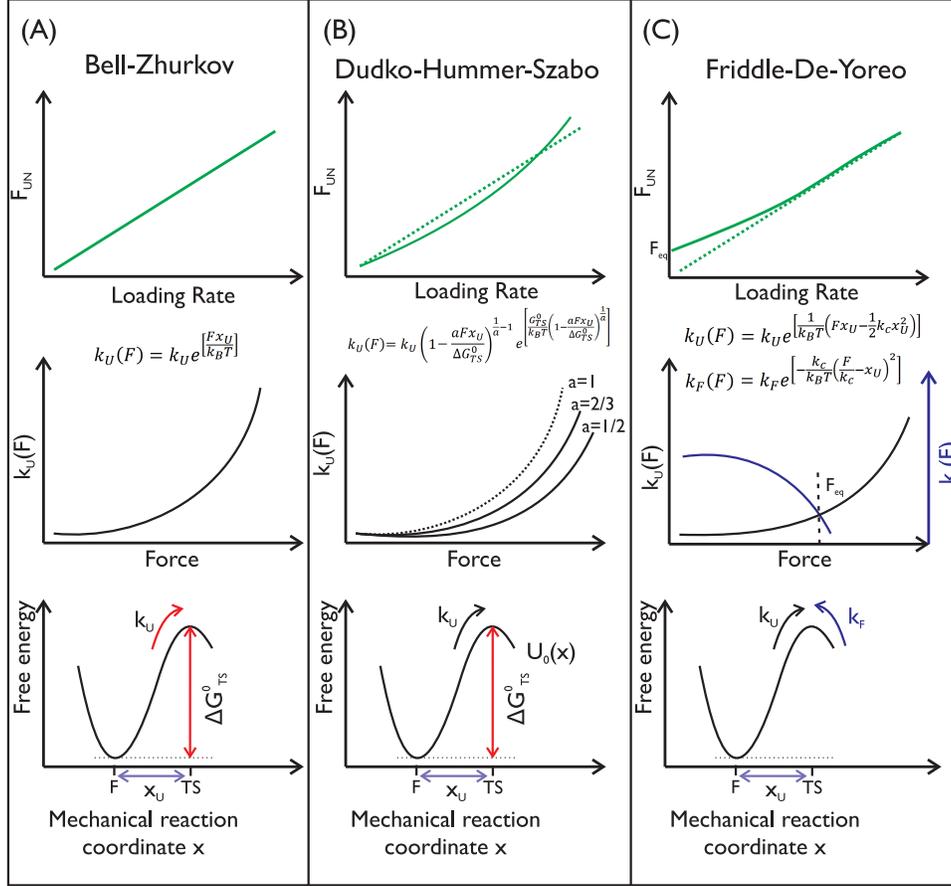


Figure 20: Schematics showing the dependence of the unfolding force on the loading rate (top), the dependence of the unfolding force on the unfolding rate, $k_U(F)$ (middle), and the energy landscape (bottom) for three models which can be applied to SMFS protein unfolding data (A) Bell-Evans-Richie[131] (B) Dudko-Hummer-Szabo[124, 133] and (C) Friddle-De-Yoreo[125]

7.2.4 Dudko-Hummer-Szabo model

In the Dudko-Hummer-Szabo (DHS) model the Bell-Evans-Richie model is reduced to its approximate limit using a stochastic model of a spring, and uses Kramers theory of diffusion to determine kinetic constants[124, 133]. The DHS model allows for differences in the shape of the energy landscape to be determined. The Bell-Evans-Richie model assumes that the unfolding force F_U scales linearly with the loading rate (Figure 20A). However, this assumption has been shown to be an oversimplification for some systems. The DHS model aims

to account for the non-linearity in the data observed in the pulling speed dependence plots of mechanical stability (Figure 20B). The theory behind the model was originally completed by Hummer and Szabo (2003) [134]. In this model the lifetime of a bond is dependent on an arbitrary exponent which is responsible for the shape of the energy landscape. Equation 28 [124, 133] gives the relationship between the change in free energy ΔG_{TS}^0 , the distance between the unfolded and transition state, x_U and the loading force, F . τ_0 is the lifetime of the bond at zero force.

$$\tau(F) = \tau_0 \left(1 - \frac{aFx_U}{\Delta G_{TS}^0} \right)^{1-\frac{1}{a}} e^{-\frac{\Delta G_{TS}^0}{e}} \left[1 - \left(\frac{aFx_U}{\Delta G_{TS}^0} \right)^{\frac{1}{a}} \right] \quad (28)$$

The equation governing this relationship can be reduced to the widely utilised Bell-Evans-Richie's model with an arbitrary exponent, $a=1$. The loading rate during an experiment, for a single molecule, can be approximated by $R = k_c v$, however, when multiple proteins or molecular linkers are present the spring constant must be modified. For the Bell-Evans-Richie model this is done by accounting for the change in effective spring constant of the protein chain in Monte Carlo simulations. Dudko et al. (2008) showed that, when molecular linkers are present, the rate needs to be corrected to Equation 29. Where p is the persistence length, L_c is the contour length, k_c is the cantilever spring constant, $\beta = (k_B T)^{-1}$ and v is the pulling speed of the experiment.

$$R(F_U) = v \left[\frac{1}{k_c} + \frac{2\beta L_c p (1 + \beta F_U p)}{3 + 5\beta F_U p + 8(\beta F_U p)^{\frac{5}{2}}} \right]^{-1} \quad (29)$$

The distribution of lifetimes, $P(F_U)$, is related to the lifetime by Equation 30.

$$P(F_U) = \frac{\exp\left(-\int_0^{F_U} [r(F_U) \tau(F)] df\right)}{R(F_U) \tau(F_U)} \quad (30)$$

Dudko et al. (2008) showed that this could be inverted to the lifetime by Equation 31.

$$\tau(F_U) = \int_{F_U}^{\infty} \frac{p(F_U) df}{[R(F_U) p(F_U)]} \quad (31)$$

This expression suggests that the distribution of the forces can readily describe the rate of unfolding of the protein, by approximating the probability of the distribution as a Gaussian distribution the expression can be approximated by Equation 32.

$$\tau(F_U) = \left[\frac{\pi}{2} \left(\langle F_U^2 \rangle - \langle F_U \rangle^2 \right) \right]^{1/2} / R(\langle F_U \rangle) \quad (32)$$

The force distributions from SMFS experiments can then be used to determine the lifetime of the bond as a function of loading rate (Equation 32), which in turn can be fitted by Equation 28. This method has the advantage of not requiring approximations of the shape of the energy landscape. However, in this form, it cannot be used for multiple barriers to unfolding, where the barriers are codependent on the preceding and following unfolding events. Zhang and Dudko (2013) have since developed this model so that it could be used for such a system[135]. In this paper they describe a method for mapping the forces from multiple unfolding events onto force dependent rates. This paper assumes that the free energy landscape has an arbitrary number of barriers with metastable intermediates between the folded and unfolded state. The population in the system that will move from state i to state j when the force changes by dF (or similarly between time t and dt) is given by:

$$P_{ij} |dF| = P_{ij}(t) dt = k_{ij}(F(t)) \mathcal{N}_i(t) dt \quad (33)$$

Where $\mathcal{N}_i(t)$ is the population of state i at time t and is dependent on the original population of this state (N_i) and the number that transitioned to state i from/to the m^{th} state. $k_{ij}(F)$ is the force dependent rate of moving between the two states. This led the authors to the following relationship:

$$k_{ij}(F) = \left| \dot{F}(F) \right| \frac{P_{ij}}{N_i(F)} \quad (34)$$

$$N_i(F) = N_i(F(t)) = N_i^0 - \sum_m \int_0^t P_{im}(t') dt' + \sum_m \int_0^t P_{mi}(t') dt' \quad (35)$$

Equations 34 and 35 can be used to map the experimental force to force dependent loading rates. Where P_{im} is the population that moved between i to m and P_{mi} is the population that moved between m to i state. $P_{ij}(F) = n_{bin}/w_{bin}$ where n_{bin} is the number of counts in a bin and w_{bin} is the width of each bin. $\left| \dot{F}(F) \right|_i$ is the loading rate in state i to correct for the constantly changing loading rate between systems. The loading rate can be read directly from the force-extension profiles as the gradient of the force-time trajectory. There are two methods for determining $N_i(F)$: by counting the number of force trajectories in state i at force F or by approximating it as the integrals in Equation 35 by the corresponding bins in the histogram. The force-dependent rates for each unfolding event could then be fit by Equation 32.

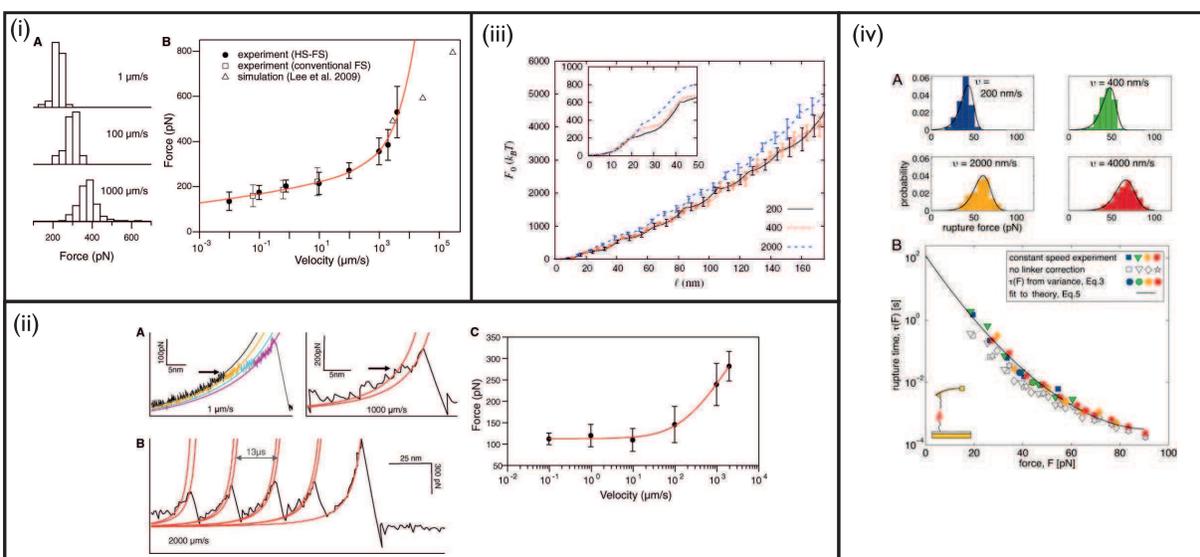


Figure 21: Examples of alternative models to the Bell-Evans-Richie model being used for force-extension data for proteins. For more information about the parameters obtained please refer to the refereed papers. Data shown in A and B is for the peak unfolding force (A) and the initial unfolding event (B) for the I27 protein. At high speed the forces deviate from a linear response predicted for the Bell-Evans-Richie model. In (A) a Hummer Szabo model [134] was used to extract the parameters and in (B) the Friddle-De-Yoreo model was employed [125]. (C) The model free approach using the modified Jarzynski equality [136] has been used to obtain the free energy profiles for a polyprotein containing 8 repeats of the I27 protein. At low velocities (red and green) the energy profiles overlay well. At higher (blue) velocities the data begins to deviate. (D) The Dudko-Hummer-Szabo model [124] was used to extract information about the underlying energy landscape for the protein ddFLN4. (A) and (B) From [119]. Reprinted with permission from AAAS. (C) Taken from [137] (D) Taken from [124] Copyright (2008) National Academy of Sciences, USA.

7.3 Thermodynamic parameters of protein mechanical unfolding

The methods described above use model-approximations to determine key features of landscapes. However, these rely upon assumptions and approximations of the underlying energy landscape, therefore there is much interest in developing and using model-free approaches [136, 138, 139] or by extracting free energy data directly from force-extension curves [108, 140]. A number of methods, based on theories by Jarzynski

[141] and Crooks [142], have been developed to determine the free energy differences from non-equilibrium experiments such as SMFS. These methods will be discussed in this section.

7.3.1 Jarzynski equality

Jarzynski [141] showed that, if a system is perturbed from equilibrium over a defined reaction coordinate by an external work, W , for a finite amount of time, then the free energy of the equilibrium state can be inferred from a series of trajectories from this initial state over the reaction coordinate. That is, if a state is driven from equilibrium the free energy of the initial state can be determined from:

$$\exp \frac{\Delta G_U^0}{k_B T} = \left\langle \exp \frac{-\beta W_\lambda}{k_B T} \right\rangle_N \quad (36)$$

where the averages $\langle \dots \rangle$ are averages of the Boltzmann factor ($\exp \frac{-\beta W_\lambda}{k_B T}$) of the external work done on the system taken from the trajectories, λ propagating from equilibrium with a time dependent Hamiltonian. Each λ has a work W_λ associated with it. This is known as the Jarzynski equality and deals with changes in the system state rather than molecular positions. This could therefore not be used directly for the data obtained from SMFS experiments which monitors the change in position of a cantilever as a function of time. Hummer and Szabo later expanded this work to generalise it for use in SMFS experiments for histograms weighted by the non-equilibrium work [136, 143] and more recently using a Weierstrass transform [138]. Using this approach, a spring is attached to the biomolecule and its extension is followed as it is perturbed by a constant velocity. The Hamiltonian is then:

$$H(x, t) = H_0(x) + V[q(t), t] \quad (37)$$

If the system is assumed to be a simple harmonic spring then $V(z(x, t), t) = k_c(z - vt)^2$ Where z is the distance between the cantilever and the surface, including the deflection of the cantilever. In such a system the Jarzynski equality becomes:

$$\exp(-\beta \Delta G_U(q)) = \left\langle \delta[z - z(t)] \exp -\beta \left(\int_C F(z, t) dz - V[z(0), 0] \right) \right\rangle \quad (38)$$

for initial points drawn from a Boltzmann distribution, where δ signifies the Dirac delta function. The $\int_C F dz$ is the integral of the force over the contour of the reaction coordinate $q(t)$ and $F(z, t) = -\partial(V(z, t))/\partial z$ and the integral is performed along the contour of the trajectory (change in position with time) from $z(t=0)$ to $z(t)$. The term $V[z(0), 0]$ represents the potential of the initial state. This is the work stored in the deflected cantilever at $t=0$ which can be determined by the zero-position of the cantilever far from the surface. For the computation of the free energy, several force-extension or force-time trajectories are required. The free energies can either be computed using a weighted histogram approach or by estimating the work done at a certain extension as the area under force-extension traces [136, 143]. These approaches are outlined in detail in references [136, 143–145]. This expression has been used to describe the free energy profiles of short RNA segments [146] and DNA hairpins [147].

However, the results for protein unfolding experiments are varied. The free energies obtained from Equation 38 should be independent of the pulling velocity used in experiments. For a number of experiments on proteins, this has not been the case [148–150] although other studies [137, 151] yielded results independent of velocities up to 1000 nm/s. It is hypothesised that these discrepancies result from the complexity of the unfolded micro-states within the “folded” state of the protein [148]. In these systems, free energies were computed in the limit of zero pulling velocity. This technique is also sensitive to the occurrences of rare events far from equilibrium which increases the uncertainties associated with the computed free energies [152].

7.3.2 Crooks fluctuation theorem

An alternative method to determine the unperturbed free energy landscape can be performed if the unfolding and refolding trajectories of proteins are obtained. This uses Crook’s fluctuation theorem which describes the symmetry of work fluctuations around forward and reverse cycles of a system driven far from equilibrium

by an external work [142]. This determines the hysteresis between the irreversible work required for the forward and reverse reactions. By performing repeated cycles of unfolding and refolding of a biomolecule performed at the same velocity, the probability distributions of unfolding, $p_u(W)$ and refolding $p_f(W)$ can be determined and related to the free energy by:

$$\frac{p_U(W)}{p_f(-W)} = \exp\left(\frac{W - \Delta G_U^0}{k_B T}\right) \quad (39)$$

Where the work can be approximated by the integral of force over the extension (i.e, $W = \int F dz$ or $W = \sum_{i=1}^N F_i \Delta z_i$, where N is the number of intervals used in the sum [153]). This states that the ratio of the two probability distributions depends only on ΔG_U^0 . At some W these two distributions will cross and $p_u W = p_f - W$. At this point $W = \Delta G_U^0$ independent of the pulling velocity. Whilst the ΔG_U^0 can be approximated by this relationship, it is more accurate to determine ΔG_U^0 from the full probability distributions. The slope from a plot of $\ln\left(\frac{p_U(W)}{p_f(-W)}\right)$ against W can then be used to obtain ΔG_U^0 . This relationship has been successfully used to reconstruct the free energy landscape of RNA [153] and polyproteins [154] but is restricted to experiments that combine unfolding and refolding measurements.

8 Protein mechanical stability

8.1 Effect of unfolding history on measured protein mechanical stability

When each individual force corresponding to a single protein unfolding event is recorded (i.e, F_U for peaks 1, 2, 3, 4, 5 in Figure 22 A) for a number of traces at a particular pulling velocity, a minimum is observed in the graph of modal force against event number (Figure 22 B). Therefore the F_U of a protein is dependent on the number of folded and unfolded proteins in a polyprotein chain [155] or the “unfolding history” of the polyprotein [155]. The minima in the forces occurs due to the competition between two effects. Firstly, the number of protein domains that remain folded in the polyprotein chain and secondly, the compliance of the polyprotein chain which depends on the cantilever spring constant and the polyprotein domain. Figure 22 (B) illustrates these effects which individually influence the loading rate applied to each folded protein domain in different ways [155, 156]. As the number of unfolded domains increases, the probability of a protein unfolding at a given force decreases. This suggests that the unfolding force would increase with event number. However, as more protein domains unfold, the length, and therefore, compliance of the chain increases. The increase in compliance decreases the effective spring constant of the polyprotein and therefore decreases the force required to unfold a protein domain.

8.2 Relationship between protein structure and mechanical stability

Over the past two decades, the force versus extension behaviour of a large number of proteins have been studied using SMFS and their mechanical stability F_U obtained. These studies have demonstrated that proteins can broadly be ranked according to their secondary structure content and arrangement – where mostly α -helical proteins are mechanically weaker (low F_U) than those predominantly composed of β -sheet structures (higher F_U). A recent comprehensive list of proteins and relevant references can be found in [78]. A selection of proteins and their measured F_U are listed in Table 1. It can be seen that while some proteins, such as I27 (the 27th immunoglobulin-like domain of titin) is able to withstand forces of above 200 pN, other proteins, such as calmodulin unfold at forces below the measurable limit of an AFM which is ~ 15 pN [8, 157]. Of the proteins studied to date, the most mechanically stable proteins are those from the all- β SCOP class, followed by α/β structures and the most mechanically labile proteins having all- α structure [82, 132, 158]. This appears to be a general trend across all proteins that have been mechanically characterised, when the N- and C-termini are aligned with the direction of the applied force.

8.3 Importance of the direction of the applied force

Despite the general trend suggesting secondary structural elements are a key determinant of protein mechanical stability, it alone cannot explain the broad range of force responses observed for proteins with similar

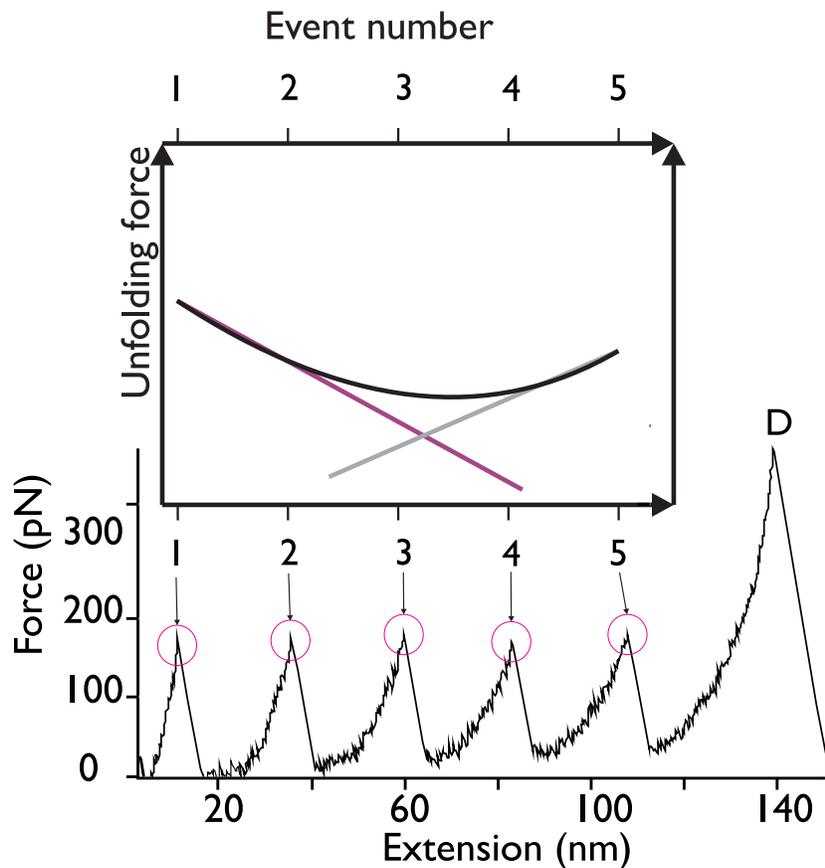


Figure 22: A force extension trace containing the unfolding events for an entire polyprotein containing five I27 domains. A single unfolding event within the polyprotein is numbered 1 to 5, known as the unfolding event number, and the detachment of the polyprotein from the tip or surface is labelled as D. Here 1 indicates the first unfolding protein event observed and 5 corresponds to the fifth or last unfolding event observed for the polyprotein. The schematic in the inset illustrates how the unfolding force changes with event number due to two competing effects. As protein domains unfold within the polyprotein, the number of folded domains decreases, this decreases the probability of observing an unfolding event at a given force and therefore acts to increase the observed unfolding force (grey linear line). As more protein domains unfold within the polyprotein, the compliance of the chain also increases. This acts to decrease the unfolding force of the chain due to a decrease in the effective spring constant of the chain (pink linear line). The competition between the two different effects causes a nonlinear relationship between the unfolding force and event number. This is known as the unfolding history effect on the unfolding force.

Table 1: Values, at a given velocity, for the maximum force a protein can withstand before unfolding. The type of structure was obtained from a search of the SCOP class of the protein.

Protein	F_U /pN (Velocity/nm/s)	Protein fold (SCOP)	Reference
I27	204 (400–600)	all β	[79]
I27 (C47S C63S)	172 (600)	all β	[132]
I1	127 (400-600)	all β	[159]
1FNIII	220 (600)	all β	[160]
10FNIII	74 (600)	all β	[160]
13FNIII	89 (600)	all β	[160]
C2A	60 (600)	all β	[82]
C2B	100 (500)	all β	[161]
E2Lip3	15 (700)	all β	[162]
TmCSP	78 (400)	all β	[78]
Barnase	70 (100-500)	α/β	[89]
Protein L	152 (700)	α/β	[132]
GB1	184 (400)	α/β	[163]
Barstar	<50 (400)	α/β	[164]
Top7	150 (400)	α/β	[164]
Ubiquitin	203 (300)	α/β	[165]
AcP	~ 60 (600)	α/β	[166]
Spectrin	~ 50 (80-800)	all α	[167]
Calmodulin	<15 pN (600)	all α	[82]

structures. For example, the β -sheet containing protein TNfn3 unfolds at a force of 120 pN at 600 nm/s [168, 169] which is significantly lower than the unfolding force for the structurally similar protein I27 [92]. Early SMFS experimental studies [82] and theoretical studies [170, 171] have suggested that the number and geometry of hydrogen bonds, with respect to the applied force, could explain the differences between structurally similar proteins. However, it was challenging to directly determine the effect of hydrogen bond geometry in different proteins due to differences in protein sequences and stabilities [162]. This problem was circumvented in two studies published in 2003. In the study by Carrion-Vazquez *et al.*, the protein ubiquitin was unfolded using two different pulling geometries: by pulling the protein between the N- and C-termini, or in a different direction, between an exposed lysine (lys48) and the C-terminus [165]. The ubiquitin domains pulled between the two termini had unfolding forces more than twice the magnitude of domains pulled in a different direction [165]. The differences in forces were attributed to the direction of the hydrogen bonds with respect to the applied force vector. This sensitivity to hydrogen bond geometry was also observed in the study by Brockwell *et al.* on the protein E2lip3 [162]. In this study the protein was immobilised on a surface at one of two positions, enabling hydrogen bonds to be either “peeled” or “sheared” apart. In the peeling geometry, the hydrogen bonds are parallel to the direction of the applied force. In the shearing geometry the hydrogen bonds are perpendicular to the direction of the applied force (Figure:23). Whilst a large unfolding force (177 pN at 700 nm/s) was observed for the protein with the shearing geometry, no force unfolding peak was observed for the protein in the peeling geometry [162]. This suggested the protein unfolded at a much lower force, below the noise of the instrument, and was as a consequence of the hydrogen bond geometry.

Since those early studies, a number of experimental [172–177] and theoretical [178–182], studies have been performed to determine the importance of pulling geometry on protein mechanical stability. For example, the pulling geometry has been controlled by circular permutations² to protein structures and by engineering exposed pairs of cysteine residues. The force response of the green fluorescence protein, GFP, was determined by applying force across five pairs of cysteine residues [172]. A diagram illustrating these results is shown in Figure 23 [172]. A large range of forces were observed between the different unfolding geometries, ranging from ~ 100 pN for the least stable geometry and ~ 600 pN for the most stable geometry [172]. Circular permutations on the protein DHFR resulted in a decrease of ~ 56 pN in the unfolding force compared with the

²A circular permutation results in an identical topology, but a different location of the N- and C- termini. This causes the secondary structure elements to unfold in an alternative order.

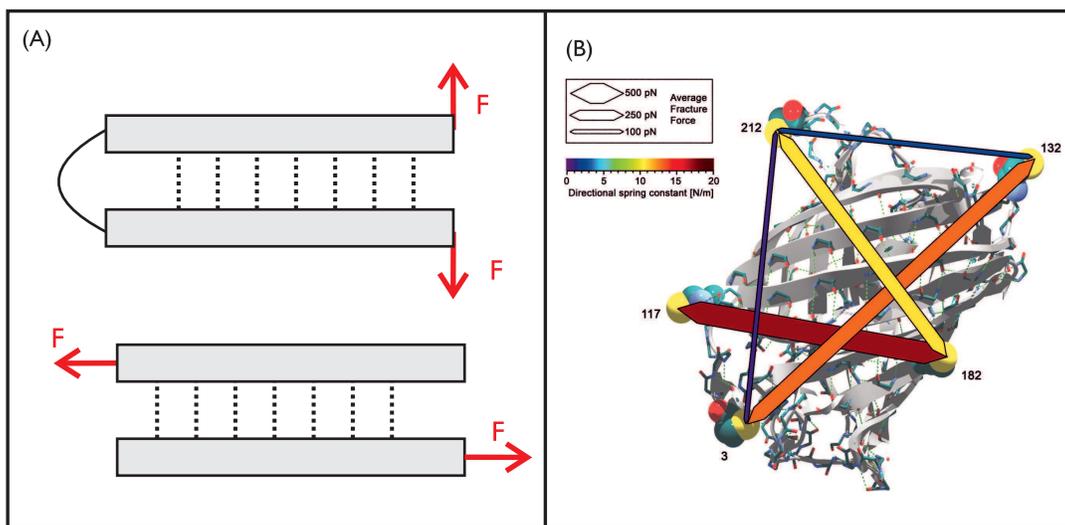


Figure 23: Illustration of the different pulling geometries of hydrogen bonded β -strands on the application of an unfolding force using SMFS. (A) Top: The “peeling” geometry where the force is applied parallel to the hydrogen bond geometry. Here the force (red arrows) is applied to the N- and C- termini of the protein such that the hydrogen bonds are broken sequentially. Bottom: The “sheer” geometry the application force perpendicular to the the hydrogen bond geometry causes the hydrogen bonds to be broken simultaneously. (B) The importance of the direction of the applied force on the mechanical stability of the green fluorescence protein, GFP, was determined by applying force across five pairs of cysteine residues [172]. Figure and values adapted from H. Dietz, F. Berkemeier, M. Bertz, and M. Rief, Anisotropic deformation response of single protein molecules, PNAS, 103 (34) 12724-12728. Copyright 2006 National Academy of Sciences, USA.

wild type protein [183]. Alternative unfolding pathways have also been observed by manipulating the pulling geometry of a protein [184]. These studies have identified key structural motifs involving hydrogen bonds that are responsible for the unfolding pathway and therefore force response of the protein. Different regions that can withstand mechanical force have also been identified for membrane proteins. Whereas globular proteins typically have a single rupture event associated with their mechanical unfolding, several mechanical barriers can exist for a single membrane protein resulting in many unfolding peaks in force-extension traces which can be related to specific structural regions, held together by hydrogen bonds and other interactions, within the protein [140] as well as the role of the membrane in biasing the folding pathway [185, 186]. Furthermore it has been shown that binding of ligands to proteins can alter the position of these mechanical barriers along the protein [187]. More recently, AFM and steered molecular dynamics (SMD) simulations have been combined to reveal force propagation pathways through a mechanically ultrastable multidomain cellulosome protein complex [188]. In this study a new combination of network-based correlation analysis supported by AFM directional pulling experiments was employed, which allowed visualisation of stiff paths through the protein complex along which force is transmitted. The results implicate specific force-propagation routes nonparallel to the pulling axis that are advantageous for achieving high dissociation forces.

8.4 Mechanical clamps motifs in proteins

The importance of the arrangement of the secondary structure in relation to the direction of the pulling force has now been demonstrated for a number of proteins, where for example the shearing apart of two β -strands requires a greater force than “un-zipping” them sequentially. An early molecular dynamics study on the I27 protein identified a ‘mechanical clamp’ region within the secondary structure which involved two neighbouring beta-strands [189, 190] (Figure 24A). A mechanical clamp is defined as a structural region in a protein that is responsible for the enhanced resistance to stretching. This element therefore confers mechanical robustness and provides the rate-limiting step for the unfolding of a protein [191–193]. In the case of the I27 protein [189], the force bearing region was identified as a cluster of six hydrogen bonds and, for the protein to unfold, simultaneous rupturing of these hydrogen bonds was required (Figure 24B). The existence of this mechanical clamp was experimentally verified in 2000 by Li *et al.* who mutated residues identified in the mechanical clamp region [194] (Figure 24C). Since then other mechanical regions in other proteins have been identified experimentally [8, 78, 132, 195, 196]. A computational study by Lu *et al.* (1999) compared the unfolding of 10 proteins containing a similar number of amino acids but with different secondary structures [170]. Using SMD, a constant velocity was applied to the ends of the proteins and the force response was monitored. This study found that elongation of proteins containing β -sheets resulted in a force peak which corresponded to the breaking of hydrogen bonds in the backbone of the protein. For proteins containing α -helices, there were no inter-strand hydrogen bonds to break in this structure and the protein unfolded readily with no force response. This suggested the hydrogen bonds connecting the β -strands in a β -sheet conformation provide more stability than hydrophobic contacts between α -helical proteins. More recently, a simple elastic model has also been used to describe this general trend between unfolding force and type of secondary structure [197]. An extensive, systematic theoretical study of protein secondary structures from the PDB permitted the identification of a number of mechanical clamp motifs [192, 193] (Figure 25). These motifs were defined according to the hydrogen bond arrangements between secondary structure elements within the protein, and have since been found to occur in many proteins across all branches of life.

8.5 Protein engineering to tune mechanical stability

Mutational studies through protein engineering have provided powerful tools to decipher the contribution of specific non-covalent interactions on protein mechanical stability. These studies have illustrated the importance of hydrogen bonded regions within proteins [194], identified intermediates in the unfolding pathways [206], highlighted the importance of hydrophobic interactions [90, 195] and identified multiple unfolding pathways within proteins [207]. As described above, hydrogen bonded motifs in the mechanical clamp region of a protein are important in determining protein mechanical stability. Mutations of amino acids within this mechanical clamp region of proteins have been shown to alter the mechanical stability of proteins [90, 194, 195, 206]. For example, the mechanical stability of the I27 protein was severely destabilised by

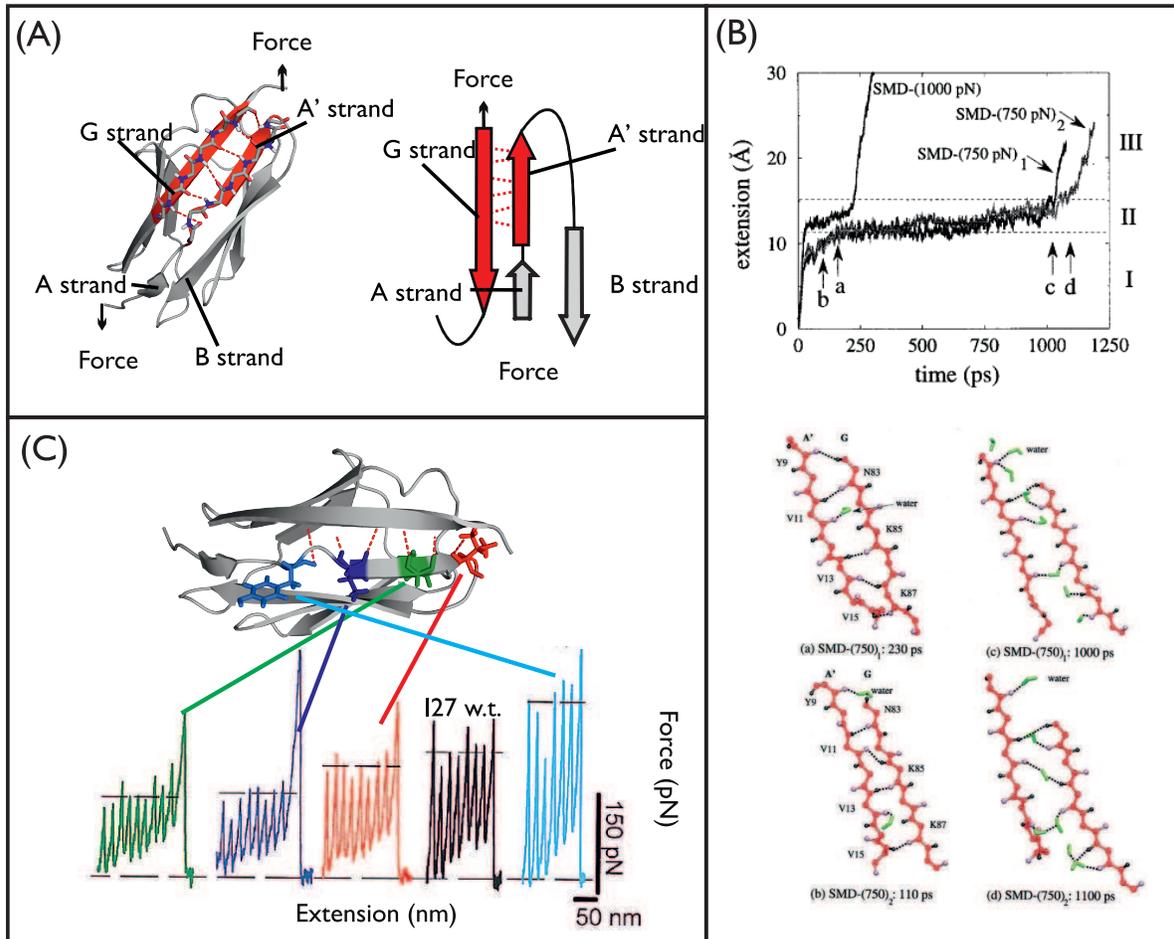


Figure 24: (A) A molecular dynamics study on the I27 protein identified a ‘mechanical clamp’ region within the secondary structure which involved two neighbouring β -strands A’ and G. The force bearing region was identified as a cluster of six hydrogen bonds and, for the protein to unfold, simultaneous rupturing of these hydrogen bonds was required [189, 190]. Molecular dynamics simulations highlighted the importance of water molecules in the breaking and reformation of hydrogen bonds in the mechanical clamp region of the I27 protein. (C) Mechanical clamp motifs have been investigated through mutating residues identified in the mechanical clamp region. (B) Reprinted from *Biophysical Journal*, 79, H. Lu and K. Schulten., The key event in force-induced unfolding of titin’s immunoglobulin domains., 5165, Copyright 2000., with permission from Elsevier. (C) Reprinted by permission from Macmillan Publishers Ltd: *Nature Structural Biology*, [194], copyright 2000.

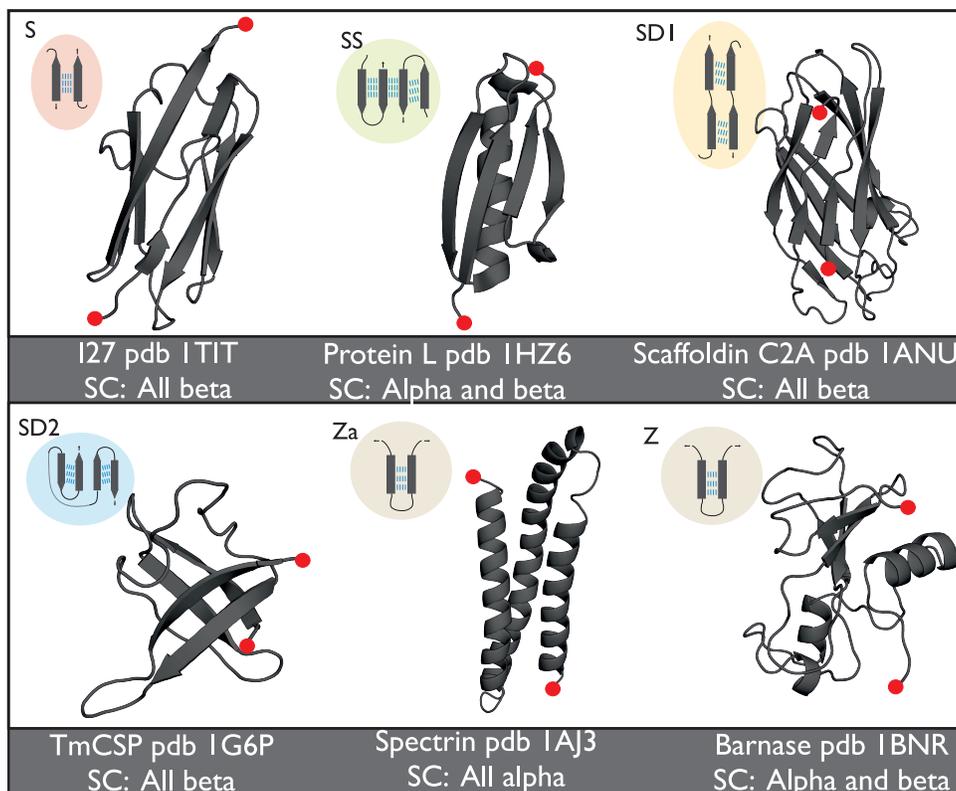


Figure 25: Figure showing tertiary structures containing α -helices and β -strands for 6 different proteins that have been mechanically characterised. The structures depicted were created using the software PyMOL [198]. The force is applied between the two termini of the protein indicated by the red dots. Each of these protein contain a mechanical clamp region, which is the force bearing region within the protein. The mechanical clamp patterns are shown to the left of the protein and satisfy patterns of mechanical clamps defined by Sikora *et al.* [192, 193]. Structures were obtained from the PDB database [199] and the references are as follows: 1TIT [200], 1HZ6 [201], 1ANU [202], 1G6P [203], 1AJ3 [204] and 1BNR [205].

mutations of hydrogen bonded amino acids within the mechanical clamp but not altered significantly by changes to the protein hydrophobic core [208]. Conversely, the hydrophobic core has been shown to play a key role in the mechanical stability of two other proteins, protein L and GB1[90, 195]. Mutational studies combined with SMFS identified a “mechanical rheostat” in the core of protein L [90]. A mutation of a single residue resulted in a 36 pN decrease in unfolding force compared with the wild type protein, despite no change in the thermodynamic stability of the protein[90]. Two other single mutations were performed on protein L, these mutations were hypothesised to increase the hydrophobic core packing. Whilst one mutation increased the unfolding force by 13 pN, the other mutation increased the force by 72 pN at a single velocity [90]. This increase in force was coupled with an increase in the thermodynamic stability of the protein. The importance of hydrophobic core packing on the mechanical unfolding force of GB1 has also been shown by mutations [195]. Protein GB1 is structurally similar to protein L. A series of mutations to this protein identified three mutations that caused a significant decrease in unfolding force. These mutations were positioned in the mechanical shearing region between two structural motifs containing hydrophobic contacts. These amino acids were hypothesised to provide stabilising hydrophobic interactions and mutating these residues disrupts the contacts between the shearing region [195].

8.6 Role of solvent interactions on protein mechanical stability

The solvent environment is important for the structure and dynamics of proteins [209–212]. In recent years there have been a large number of studies to understand the role of osmolytes on protein mechanical stability. Osmolytes are small organic molecules that can shift the equilibrium between the folded and unfolded states of a protein [213]. So called protecting osmolytes, such as glycerol, sorbitol and sucrose, push the equilibrium towards the folded state of the protein. Denaturing osmolytes such as guanidine hydrochloride (GdnHCl) and urea shift the equilibrium to the unfolded state of the protein. SMFS studies have been performed in solvent environments containing a number of different osmolytes [214–216]. For example, glycerol has been shown to increase the mechanical stability of the protein GB1 [216], while denaturing osmolytes such as GdnHCl and urea have been found to decrease the mechanical stability of proteins[215, 217]. The mechanical properties of a protein can be tuned by a change in pH or ionic strength of a buffer [81, 81, 218]. Electrostatic interactions within a protein can be affected by changes in the environment. Tuning the pH of a solvent has been shown to “switch-on” or “switch-off” electrostatic attractions and repulsions within a protein [81]. Furthermore, these electrostatic interactions can be screened by the introduction of salt [81, 219]. More recently, SMFS has also been used to mechanically probe a protein in a nonpolar environment [220]. Moving the protein from a water-based solvent to an organic nonpolar solvent octyl benzene, resulted in the loss of mechanical stability of the protein. This was proposed as being equivalent to unfolding the protein in a high (6 M) concentration of GdnHCl and therefore suggested that the mechanical resistance of the protein was lost due to the loss of the tertiary structure of the protein[220].

8.7 Temperature softening of protein mechanical properties

Mechanical flexibility is important for the function of many proteins. SMFS has been employed to measure the stiffness of single proteins as a function of temperature. In an early study by Rief *et al.* SMFS experiments were completed on a single domain of *Dictyostelium discoideum filamin* (ddFLN4) in a temperature range from 5°C to 37 °C[221]. Force extension experiments were completed and the pulling speed dependence of the F_U was used in combination with the Bell-Evan-Richie model to determine parameters which described the unfolding energy landscape: the transition barrier heights (ΔG_{TS}^0) and position (x_U). This study found a narrowing of unfolding force distributions with increasing temperature, and an increase in the x_U from 2.7 Å to 7.8 Å. From the unfolding energy landscape parameters it is possible to estimate the spring constant of the protein along the direction of pulling. Assuming a parabolic function for the free energy $G(x) = 0.5Dx_U$, where D is the spring constant, they measured a reduction of the spring constant by a factor of 7 upon heating across this temperature range. The study suggested this temperature softening reflected a shift in the nature of the interactions responsible for mechanical stability from hydrogen bonds to hydrophobic interactions. The effect of temperature on the mechanical stability of the I27 has also been determined using SMFS. Again, the unfolding force distributions were found to narrow significantly as the temperature

increased, suggesting that the x_U between the folded and transition states in the energy landscape along the pulling direction is increased. This temperature softening has more recently been observed in a cold shock protein from the hyperthermophilic organism *Thermatoga Maritima* [91].

8.8 Measuring energy landscape roughness of proteins using mechanical unfolding experiments

An alternative explanation for the effects of temperature noted above are that SMFS experiments of protein unfolding can be used to extract information about disorder and roughness, η of the underlying energy landscape[121, 222, 223]. Length-scale dependent ruggedness in the free energy landscape of a protein arises due to topological frustration[223]. For a protein which (un)folds in a two-state manner, the roughness is thought to arise from the presence of multiple energetic interactions, and is likely to be small. Although no direct measurement of η has been made for biomolecules, it has been inferred by using a diffusion limited time scale for loop closure in model peptides, that $\eta/k_B T \approx 2$ [224]. If the unfolding energy barrier is assumed to be temperature-independent, it is possible to calculate the roughness of the unfolding energy landscape from the variation in x_U with temperature[225]. This approach has been applied to measure the η in a number of proteins, yielding values of $\eta \sim 4.4 k_B T$ for the cold shock protein[91], $\eta \sim 4.0 k_B T$ for ddFLN4[221] and $\eta \sim 4.3 k_B T$ for I27[226].

8.9 Mechanical Phi-value analysis probes mechanical unfolding transition state

Mutational analysis is a powerful tool that uses minimal perturbations of the system to interrogate the molecular origin of protein kinetic stability. A mechanical Φ -value analysis method allows a detailed examination of the structure of the transition state in a force unfolding reaction[227]. This is the mechanical equivalent of the well-established thermodynamic analysis which is used to examine the effect of a mutation based on the relative changes in the free energy of the native state, transition state and denatured state of a protein[228]. The mechanical Φ -value is determined by comparing the effect of mutation on the transition state and measures the amount of native structure present around the mutated residue in the transition state. This comparison enables information about the native contacts formed at the transition state upon folding in the mutated protein to be obtained[227]. The mechanical Φ -value is given in equation 40 [227] and is defined as the ratio of the change in free energy of activation for folding ($\Delta\Delta G_{TS}^0$), to the change in equilibrium free energy ($\Delta\Delta G_U^0$).

$$\Phi = 1 - \frac{\Delta\Delta G_{TS}^0}{\Delta\Delta G_U^0} \quad (40)$$

A value of 1 indicates that the protein is fully native at the transition state. A partial value in SMFS experiments indicates partial structuring of the protein [227]. Φ -value analysis has been used to determine the role of different amino acids on the transition state of the protein. Best *et al.* determined the regions of the protein that were involved in the mechanical unfolding event in the protein I27 [227]. Furthermore, the Φ -values obtained for the destabilising hydrophobic deletions in both Protein L and GB1 indicated that certain mutated amino acids were only partially structured at the transition state [90, 195]. This sensitivity of SMFS experiments to the structure of the single molecules at the transition state enables determination of important interactions at the rate limiting step.

9 Towards design rules for predicting the mechanical properties of proteins

9.1 Predicting the mechanical stability and malleability of proteins

SMFS can now accurately provide access to the mechanical properties of proteins. Is it possible to develop design rules which can predict the mechanical properties of proteins? In a recent attempt to answer this question, a review was completed of the available SMFS experimental data on protein unfolding. We uncovered a number of empirical relationships between the measured mechanical stability of a protein and its

malleability, which provided a set of equations that might be employed to estimate properties of previously uncharacterised proteins. This provides a basic toolbox of correlations that permits the estimation of three important parameters of a protein, the unfolding force at an unmeasured pulling speed, the distance to the transition state and the unfolding rate at force zero (k_U). This offers an attractive, high throughput tool for identifying target proteins for desired applications where knowledge of the mechanical properties are required in a timely and accurate manner without the need for time-intensive experiments. This reveals features such as a correlation between the type of mechanical clamp in a protein and its mechanical properties[229]. For example, the Δx_U and the unfolding force (Figure 26). Typically proteins with the SD1 mechanical clamp motif, using the nomenclature of Cieplak et al [191–193], have a smaller x_U and a larger range of unfolding forces. This study demonstrates that similarities in the structural topologies of proteins results in similarities in the mechanical unfolding energy landscapes. The relationship between the x_U and F has been shown to follow a power law suggesting proteins with a lower unfolding force have a higher x_U (forces determined at 600 nm/s) [229]. These studies also reveal gaps in the explored space of mechanical properties of studied proteins (Figure 26), which will be helpful for the selection of proteins for future force spectroscopy studies. Recent studies on the α/β protein, the POTRA domain 2 of BamA from *E. coli* (EcPOTRA2), along with examination of the mechanical properties of other previously characterised α/β proteins identified a clear correlation between mechanical stability and malleability (Figure 27)[80]. These α/β proteins all share a common structural feature of proximal terminal β -strands in parallel geometry. Such studies improve our understanding of the relationship between structure and mechanical strength, increasing our ability to design proteins with tailored mechanical properties.

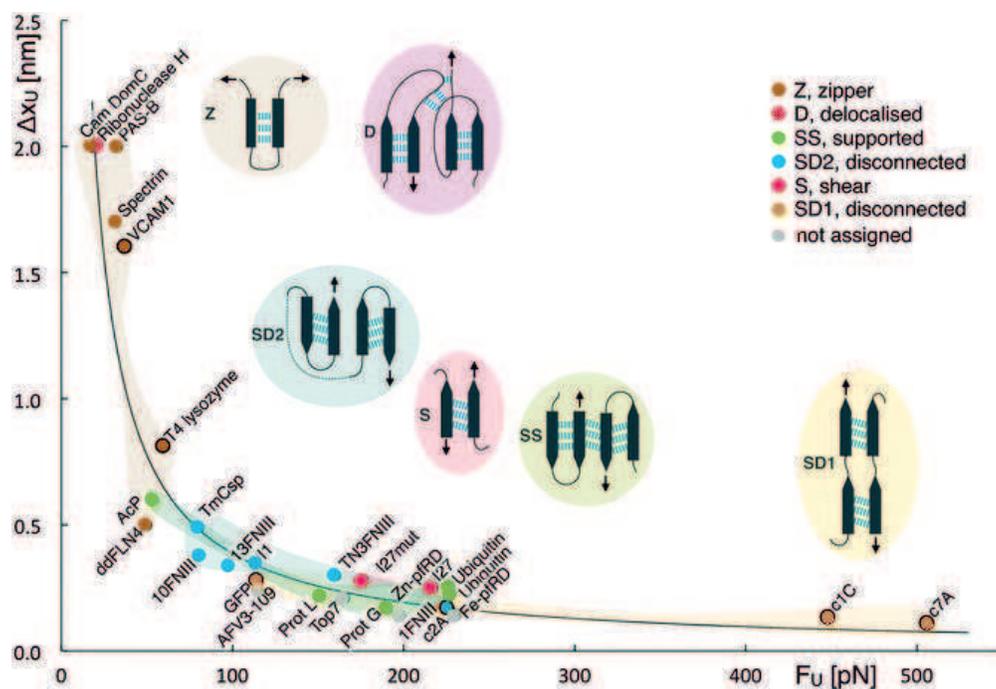


Figure 26: Examples of mechanical clamp motifs which have been identified in proteins using the notation developed by Sikora *et al.*, including (S) shear, (SA) shear antiparallel, (Z) zipper, (SD1) shear disconnected 1, (SD2) shear disconnected 2, (SS) shear supported, (T) torsion, (D) shear delocalised. The relationship between Δx_U and F_U at a pulling speed of 600 nm s⁻¹ is shown along with a non-linear fit following a power law fit. Figure taken from Reprinted with permission from [229]. Copyright 2013 American Chemical Society.

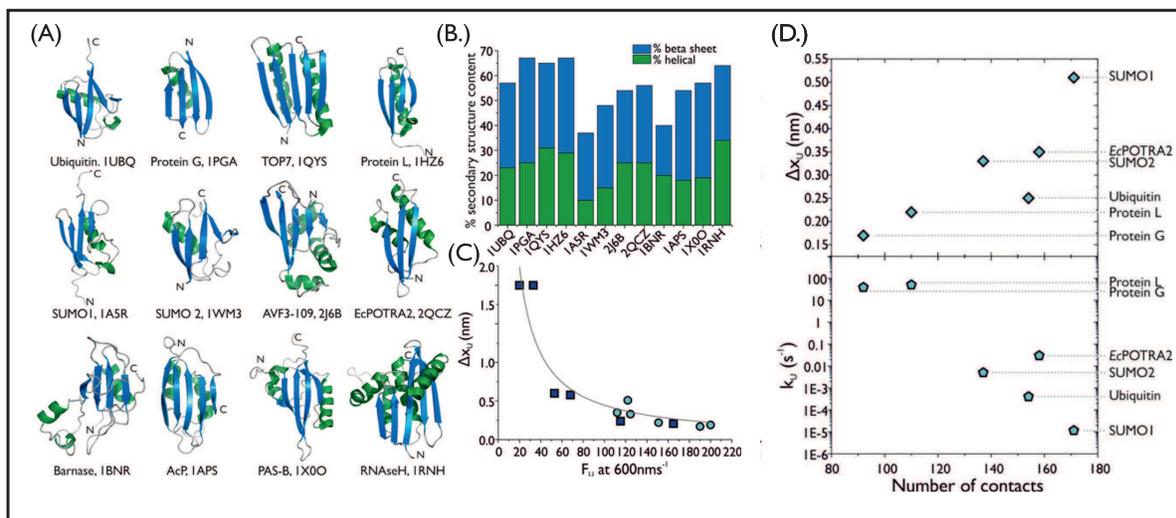


Figure 27: Mechanical stability of α/β proteins. (A) α/β proteins studied using SMFS for which a pulling speed dependence of the mechanical stability has been obtained. Proteins are shown in ribbon representations with β -strands as blue arrows and α -helices as green ribbons. The proteins are extended from their amino- and carboxy-terminal ends (N and C). The relevant PDB accession code is shown for each protein. The proteins are ordered by their unfolding force F_U at 600 nm s^{-1} , with ubiquitin having the largest F_U and RNaseH the smallest F_U . (B) Histograms show the percentage of α -helical (green) and β -sheet (blue) content in each protein, calculated from the PDB structure. The proteins are again ordered by their unfolding force F_U at 600 nm s^{-1} . (C) Graph showing the unfolding force F_U at a pulling speed of 600 nm s^{-1} and the unfolding distance Δx_U for α/β proteins which have been experimentally unfolded at least at two different speeds. Where required, the expected unfolding force at 600 nm s^{-1} was interpolated. The data can be described by a bootstrapped, nonlinear fit following a power law with $\Delta x_U = 45.0 \pm 10.5/F_U$ ($R^2 = 0.86 \pm 0.01$). Examination of the structure of the 12 α/β proteins reveals a subset of proteins (blue circles) that possess a similar structural feature: directly connected, parallel alignment of the N- and C-terminal β -strands. α/β proteins that lack this structural feature are shown as dark blue squares. (D) The number of inter-residue contacts correlates with Δx_U (upper) and with k_U (lower) for the six α/β proteins. Adapted with permission from [80]. Copyright (2015) American Chemical Society.

9.2 Predicting the critical rupture force of proteins

In SMFS we can measure the mechanical stability of proteins. Can we predict the mechanical response of a protein to force? Recently a number of models have attempted to do just this. Given hydrogen bonds play a key role in the mechanical stabilisation of proteins, a number of models have incorporated hydrogen bonding propensity in an attempt to predict the mechanical stability of proteins, and to explain differences in mechanical stability for different types of structure. For example, Gabovich *et al.* have developed a model to explain the increase in stability exhibited by β -sheet proteins in comparison with α -helical proteins [197]. In the model, the β -sheet protein is modelled by a planar zig-zag, where each rod-like section is a tube of equal length, b and sections are connected by hinges. The angle between each segment is α and is considered to be equal. The formula used to calculate the critical force for unfolding of a β -sheet is given by:

$$F_U^{eq} = (\pi d_{tube}^2 E)/4 \left[\frac{1 - \sin\left(\frac{\alpha}{2}\right)}{\sin\left(\frac{\alpha}{2}\right)} \right] \quad (41)$$

Where F_U^{eq} is the critical unfolding force at equilibrium (where the pulling velocity $v = 0$ nm/s and $T = 0$ K), $(\pi d_{tube}^2 E)/4$ is the effective cross section of the protein and d is the diameter of the tube. It must be noted that the Young's Modulus E is included in this model. The authors estimate it to be 0.2 GPa for I27 based on previous studies [230]. Interestingly, other work has shown that E can be determined from force-extension curves [231, 232]. In one study, a model could be used to extract information from the force versus extension curve and derive how the Young's modulus changes as the extension of the protein increases [232]. The unfolding force for the β -sheet protein I27 obtained using this model was $F_U^{eq} \sim 300$ pN, which is in the same order of magnitude as the force seen in experiments but is determined at $T = 0$ K and $v = 0$ nm/s. The study suggests this may be an overestimate of the F_U^{eq} . The authors partially attribute errors associated with the force to the error in the value of E used in the study. In the same study, single α -helices are modelled as mechanical springs. The spring is composed of coiled wire where the diameter of each of the n coils (pitch of the spring) is D_0 and the diameter of the wire composing the spring is d_{spring} . Under a mechanical load, the perpendicular plane to the coils makes an angle θ to the direction of the force and the current diameter of the spring, D , depends on this angle. The length of the spring under load is $L = \pi D n / \cos\theta$. As the protein fully elongates $D \rightarrow d_{spring}$ and θ will go from ~ 0 to $\pi/2$. The formula used to calculate the critical force for unfolding of a α -helix is given by

$$F_u^{eq} = \frac{3\pi G d_{spring}^2}{8 (D_0/d_{spring})} \left(\frac{(\pi D_0/d_{spring}) - 1}{4 (D_0/d_{spring})^2 + 3} \right) \quad (42)$$

Where G is the shear modulus. For α -helical proteins, the F_u^{eq} varies with the determined (D_0/d_{spring}) . The F_u^{eq} , determined for helical proteins with $1 < (D_0/d_{spring}) < 0$, was found to be between 2 – 3 pN. This suggests that the forces for α -helical domains at small pulling speeds are just a few pN, suggesting they are significantly more labile than the β -sheet proteins. This elastic theory is in agreement with experimental results on mechanical denaturation of proteins. The advantages of these models are that there are just two free parameters for both β -sheet and α -helical domains which can be estimated experimentally or from the protein structure. However, due to the simplicity of the model, there are key downfalls with it: it does not account for thermal fluctuations; the protein is assumed to be at 0 K and intermolecular bonds such as hydrophobic interactions are not quantitatively included. Despite this, it successfully accounts for the differences in mechanical properties of α -helical and all β structures of proteins.

9.3 Propagation of forces in proteins

In another recent study a 'crack propagation' model was developed to describe the critical unfolding force, F_{break} of β -sheet rich proteins [233, 235, 236]. The theory of crack propagation is based on the driving force of the crack being governed by the total potential energy of the system, which in turn is governed by the total elastic energy stored in the protein and the work done by external load on the protein. In the study, a fracture was proposed to occur when the potential energy gained by an extension from a crack balances the energy necessary to create a new surface. A crack in the context of this model is defined as

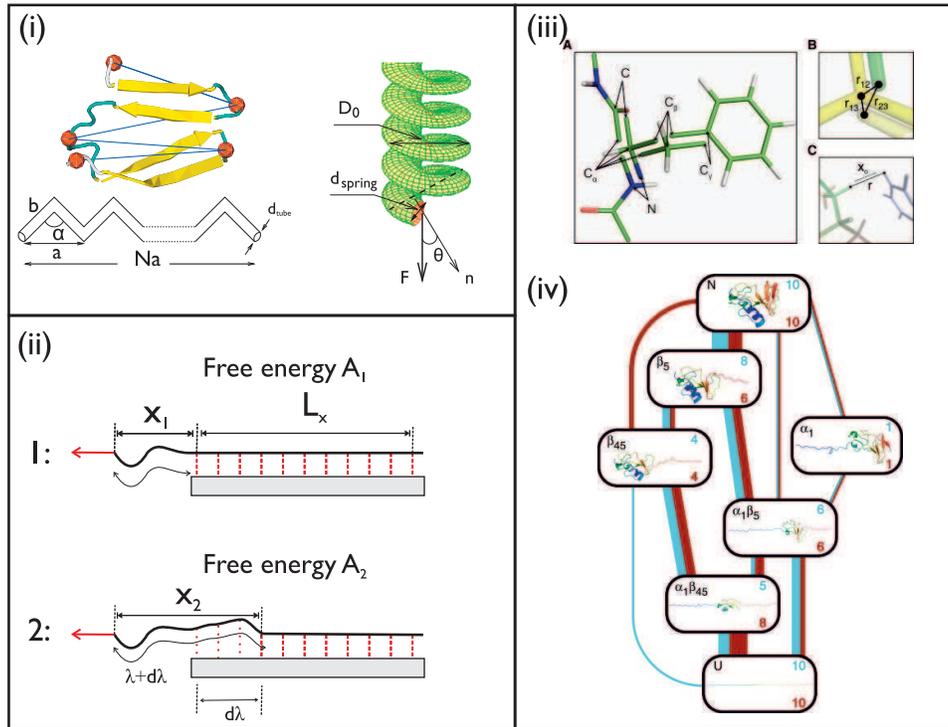


Figure 28: (A) A model has been developed by Gabovich *et al.* to demonstrate the differences in forces observed between proteins containing β -sheets and α -helices [197]. In this model the β -sheet is modelled by a “zig-zag” (right). Each strand is modelled as a tube with diameter d_{tube} with length b that make an angle α with the adjoining strand. The α -helices are modelled as a spring formed of n turns of a wire (right). The wire has a diameter d_{spring} and the pitch of the turns in the coil have a diameter of D_0 . The applied force makes an angle θ to the plane perpendicular to the coils. (B) Schematic of how fracture in a hydrogen bond network is modelled to propagate due to an applied force on a double stranded system [233]. Hydrogen bonds are formed between two strands which are perpendicular to the direction of applied force (i) with a total length L_x which is assumed to tend to ∞ . When rupture occurs (ii) the contour length (λ) of the detached strand increases by λ . This causes a change in the free energy, A . (C) Principles behind the modelling of an amino acid in a protein structure in the force-crack propagation simulation [234]. In A the different atoms comprising a single amino acid are labelled and are disjointed for clarity. B shows how the rigid units are constrained between two amino acids and C shows how steric repulsion is achieved between parts of the protein. (D) An example of how the unfolding pathways mapped for barnase using the crack propagation network (blue) and typical MD simulations (red) differ. (A) From F. Rico, L. Gonzalez, I. Casuso, M. Puig-Vidal, and S. Scheuring. High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations. *Science*, 342(6159):741–743, 2013. Reprinted with permission from AAAS. (B) Adapted with permission from S. Keten and M. J. Buehler. Geometric confinement governs the rupture strength of H-bond assemblies at a critical length scale. *Nano Lett.*, 8(2):743–748. Copyright 2008 American Chemical Society. (C) Reprinted from *Biophysical Journal*, 101/3, Graff, A. M. R. D., Shannon G., Farrell D.W., Williams P. M. and Thorpe M. F., Protein unfolding under force: crack propagation in a network, 736–744, Copyright 2011, with permission from Elsevier.

the breaking of hydrogen bonds within the structure of the protein. By coupling this idea with the force-extension relationship for polymer elasticity, which has been widely used to describe protein unfolding, the authors derived an expression for the critical unfolding force, F_{break} :

$$F_{\text{break}} = (k_B T) / (4L_p) [(1 - \alpha_{cr})^{-2} + 4\alpha_{cr} - 1] \quad (43)$$

Where L_p is the persistence length of the chain, which is a measure of the flexibility of a chain and can be determined experimentally, α_{cr} is the critical ratio between the extension and the contour length for rupture to occur. This parameter can be determined from the energy of a hydrogen bond and the length over which a hydrogen bond acts. Using this model, it has also been shown that rupture of clusters of $\sim 3-4$ hydrogen bonds are responsible for the mechanical properties of β -sheet structures which has implications for the design principles behind biological materials [233].

10 Future Directions

10.1 Lessons from extremophiles: SMFS studies of extreme-loving proteins

The non covalent forces that stabilize protein structures are still not fully understood[237]. Proteins from extremophilic organisms provide excellent model systems to determine the origin of protein thermodynamic and mechanical stability. Extremophiles are organisms which survive and thrive in extreme environments and are classified on the basis of the particular extreme environmental condition in which they live. The proteins from extremophilic single-celled organisms have received considerable attention as they are structurally stable and functionally active under extreme physical and chemical conditions, including extremes of temperature, salt, acid, and pressure, all of which have relevance in biotechnological research and industry. Much valuable work has been done to understand the structural adaptations of proteins from extremophiles. By comparing high-resolution structural data of homologous proteins from mesophilic and extremophilic organisms, a large degree of secondary structural similarity has been found, while differences in the sequences of proteins often result in large variability in thermostability. Structural observations reveal a progressive pattern of stabilisation of proteins from extremophilic organisms through multiple additional interactions at solvent exposed, loop and interfacial regions. We have recently applied SMFS to manipulate proteins from extremophilic organisms, to gain information about their stability, flexibility and their underlying energy landscapes (Figure 29). Using SMFS force extension experiments we measured the mechanical stability of the cold shock protein *TmCsp* from the hyperthermophilic organism *Thermatoga maritima* in the temperature range of 5 – 40 °C[213]. We showed that mechanical signature of *TmCsp* is maintained over the entire temperature range and observed temperature-dependent changes in features of the unfolding energy landscape, by studying the pulling speed dependence of the unfolding force with temperature in combination with Monte Carlo simulations. The main effect of temperature is seen in the distance between the native, folded state of the protein and the unfolding transition state, Δx_U which was large and increased with increased temperature, reflecting a reduction in the spring constant of the protein and an increase in the malleability or deformability of the structure. This result was surprising because malleability within the native folded basin is contrary to the hypothesis that proteins from hyperthermophiles should have rigid structures resulting from improved packing and increased numbers of ionic interactions. Instead, these experiments suggest that enhanced malleability at higher temperatures may enable proteins from hyperthermophilic organisms to maintain and easily reform their structure when exposed to denaturing high temperature conditions. This approach, of examining proteins from extremophiles, lays the foundation for further studies to examine the role of non covalent interactions in conferring protein stability and flexibility.

In a recent example, we demonstrated that a protein from a hyperthermophilic organism can be used to determine the role of salt bridge interactions in determining protein stability [238]. SMFS experiments showed that at ambient temperatures *TmCsp* is mechanically stronger yet, counter-intuitively, its native state can withstand greater deformation before unfolding (i.e. it is mechanically soft) compared with the mesophilic cold shock protein, *BsCsp*. MD simulations were used to identify the location and quantify the population of salt bridges, and revealed that *TmCsp* contains a larger number of salt bridges than *BsCsp*. To test the hypothesis that salt-bridges endow these mechanical properties on the *TmCsp*, a charged triple

mutant (CTM) variant of *BsCsp* was made by grafting an ionic cluster from *TmCsp* into the *BsCsp* scaffold. We showed that a grafted ionic cluster can increase the mechanical softness of a protein and speculate that it could provide a mechanical recovery mechanism and that it may be a design feature applicable to other proteins.

10.2 Folded protein-based biomaterials

As well as understanding the properties of proteins in isolation, there is now considerable interest in the use of proteins as components in bionanomaterials with specific and desired functions. For example, elastomeric proteins act as important functional units in biomechanical machinery. These proteins are being exploited as building blocks for biological materials that exhibit attractive and bespoke mechanical properties, as they possess the desired elasticity, mechanical strength and resilience required for these functional materials. Recent studies have demonstrated examples of engineered elastomeric proteins with mechanical properties that mimic and surpass those of natural elastomeric proteins, and have utilised natural elastomeric proteins that are well-characterised on the nano-scale to engineer hydrogels with specific macro-scale mechanical properties[240–243]. Central to these novel developments is the ability to measure the mechanical properties of the building block, the protein. As the field of protein-based biomaterials expands, there is a need to move towards predictive tools that can rationally identify proteins with desired material properties. Ideally, it would be advantageous if a protein could be selected which has the required mechanical properties, for example in terms of its mechanical stability and malleability. This would greatly expand the tool-kit of protein 'parts', given there are now over 100 000 protein structures in the PDB.

11 Conclusions

The ability to manipulate single protein molecules using SMFS and measure their force versus extension behaviour has provided exciting opportunities to understand their mechanical properties. The field has advanced, thanks to the development of instrumentation and the methods to analyse and interpret the data[245–247]. SMFS is now embedded within the scientific community and is being exploited to probe a diverse range of biological systems, including proteins, DNA, RNA and their complexes[67, 81, 85, 157, 248–252]. While early experiments provided the first measurements of protein mechanical stability, further experimental and theoretical studies have pin-pointed the molecular origin of this resilience. This has included descriptions of hydrogen-bonded mechanical clamp motifs, the role of the solvent environment and ligand binding. From this knowledge it is now possible to work towards a set of design rules or principles that might be able to predict the mechanical properties of proteins. Such tools would be powerful for the realisation of exploiting proteins as biological parts within more complex structures, in which their inherent mechanical stability might provide a useful characteristic. This could include sophisticated biomaterials with bespoke mechanical properties. Indeed, we might not be far from an experimental system in which the mechanical properties of the protein building block can be fine-tuned such that the properties and function of the biomaterial can be perfected.

12 Acknowledgements

Lorna Dougan is supported by a grant from the European Research Council (258259-EXTREME BIOPHYSICS). We wish to thank Drs David Brockwell and Matthew Batchelor for critical reading of the manuscript and for providing detailed feedback, as well as helpful discussions and comments on the manuscript from Ben Connelly and Oliver Harrison.

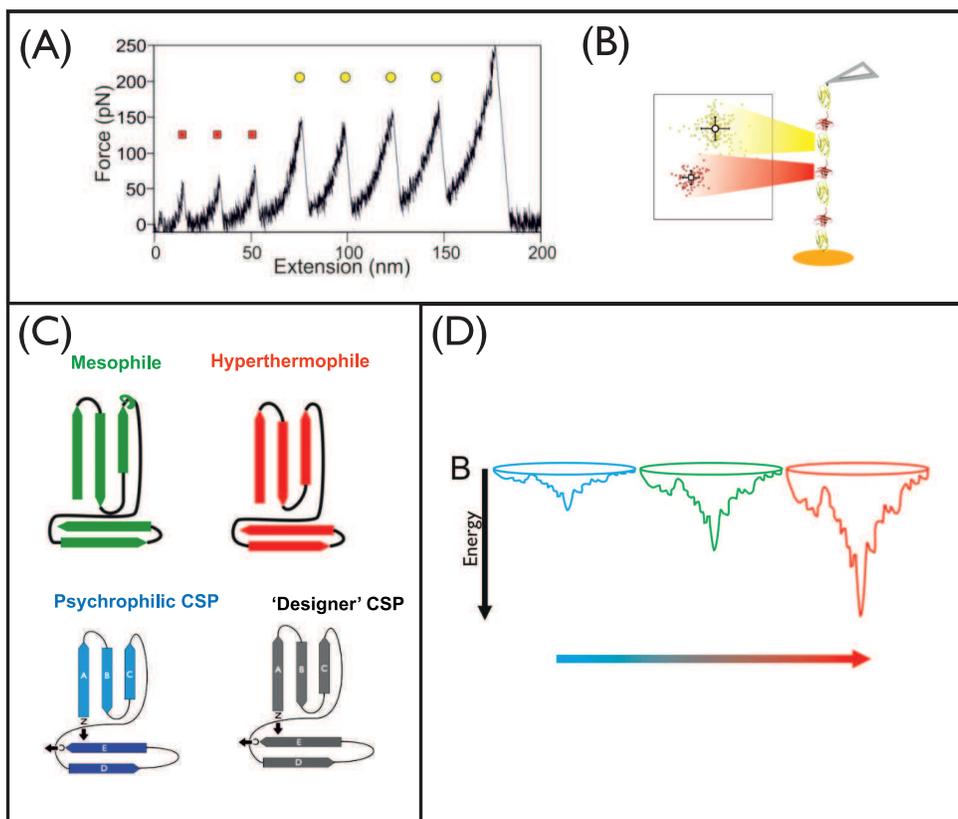


Figure 29: (A) An example AFM experimental force-extension trace showing the unfolding of a full polyprotein chain containing three *TmCsp* domains (red squares) and four I27 domains (yellow circles) at room temperature, and a pulling speed of 600nm s^{-1} . The measured peak unfolding forces (F_U) and inter-peak distances x_p for each unfolding event are recorded. (B) Scatter plot of inter-peak distances and peak unfolding forces for *TmCsp* and I27 from a single experiment at 600nm s^{-1} at room temperature. The median F_U and x_p values are plotted as symbols with black outlines. (C) Proteins from different variants of extremophile organisms, including psychrophilic, mesophilic and thermophilic, provide model systems with which to explore the origin of protein stability and provide an opportunity to determine 'design rules' for new proteins with bespoke mechanical properties and stability. (D) Schematic depicting the energy funnel model of protein folding for proteins from psychrophilic, mesophilic and thermophilic organisms [239]. The lowest free energy state corresponds to the most stable, native, folded state of the protein. The higher energy states correspond to random-coil and unfolded protein structures. Proteins from psychrophilic organisms are generally less stable than their mesophilic and thermophilic counterparts, meaning that it is easy for them to interchange between different structures. This is an essential adaptation to enable them to function at lower temperatures. The ruggedness of the bottom of the energy funnel depicts the energy barriers for inter-conversion or structural fluctuations of the native state. Figures A and B are reprinted with permission from Journal of Physical Chemistry B, 117(6): 1819-26 (2013). Copyright 2013 American Chemical Society. Figure D: This research was originally published in Biochemical Society Transactions. K. M. Tych, T. Hoffmann, M. Batchelor, M. L. Hughes, K. E. Kendrick, D. L. Walsh, M. Wilson, D. J. Brockwell, L. Dougan, Life in extreme environments: single molecule force spectroscopy as a tool to explore proteins from extremophilic organisms. Biochemical Society Transactions. 2015; Volume: 179-185 © the Biochemical Society.

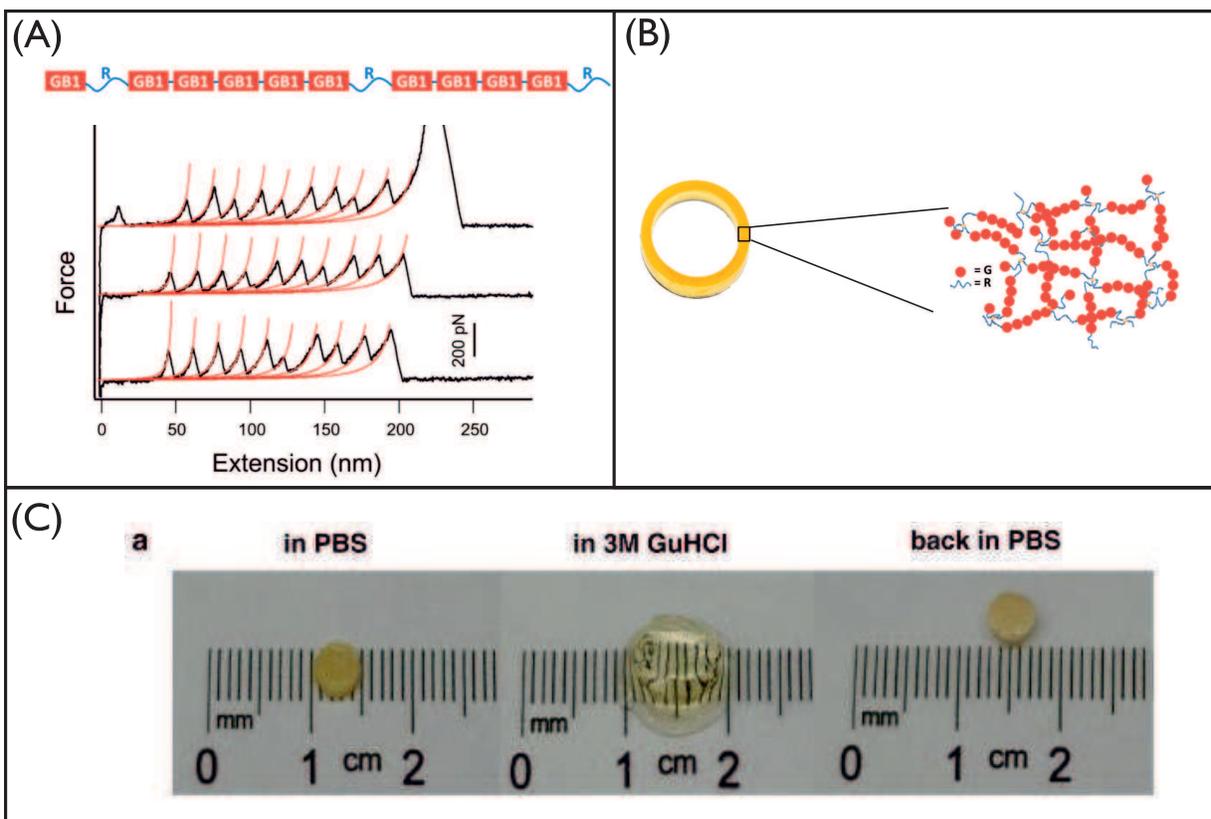


Figure 30: (A) Engineering elastomeric protein-based biomaterials via the bottom-up approach. (A) Schematic of miniature-titin-like elastomeric protein GRG5RG4R. G represents GB1 domain and R represents resilin consensus sequence. Bottom: Force-extension curves of GRG5RG4R. The initial featureless spacer corresponds to the stretching of unstructured resilin sequences, and the sawtooth peaks correspond to the unfolding of GB1 domains. (B) A photograph of a hydrogel ring constructed from GRG5RG4R and the schematic of the network structure. (C) Physical properties of hydrogel, (FL)₈, made from a de novo designed protein that assumes the classical ferredoxin-like fold structure reversibly tuned by GuHCl. In PBS, the hydrogel is opaque and shows a negative swelling ratio. Upon soaking in 3M GuHCl, the hydrogel changes to transparent and swells significantly. Figures (A) and (B) Reprinted with permission from ([244]) H. Li and Y. Cao, *Protein Mechanics: From Single Molecules to Functional Biomaterials*, *Accounts of Chem. Research*, 43(10), 1331-1341, 2010. Copyright 2010 American Chemical Society. Figure (C) Reprinted by permission from Macmillan Publishers Ltd: *Nature Communications*, ([242]), copyright 2013.

References

- [1] K. C. Neuman and A. Nagy. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods*, 5(6):491–505, 2008. doi:10.1038/nmeth.1218.
- [2] A. Ashkin, J. M. Dziedzic, and T. Yamane. Optical trapping and manipulation of single cells using infrared laser beams. *Nature*, 330(6150):769–771, 1987. doi:10.1038/330769a0.
- [3] S. B. Smith, L. Finzi, and C. Bustamante. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*, 258(5085):1122–1126, 1992. doi:10.1126/science.1439819.
- [4] L. Chen, A. Offenhäusser, and H.-J. Krause. Magnetic tweezers with high permeability electromagnets for fast actuation of magnetic beads. *Rev. Sci. Instrum.*, 86(4):044701(1)–044701(9), 2015. doi:10.1063/1.4916255.
- [5] K. Mitsui, M. Hara, and A. Ikai. Mechanical unfolding of alpha(2)-macroglobulin molecules with atomic force microscope. *FEBS Lett.*, 385(1-2):29–33, 1996. doi:10.1016/0014-5793(96)00319-5.
- [6] J. Zlatanova, S. M. Lindsay, and S. H. Leuba. Single molecule force spectroscopy in biology using the atomic force microscope. *Prog. Biophys. Mol. Biol.*, 74(1-2):37–61, 2000. doi:10.1016/S0079-6107(00)00014-6.
- [7] H. Clausen-Schaumann, M. Seitz, R. Krautbauer, and H. E. Gaub. Force spectroscopy with single bio-molecules. *Curr. Opin. Cell Biol.*, 4(5):524–530, 2000. doi:10.1016/S1367-5931(00)00126-5.
- [8] T. Hoffmann and L. Dougan. Single molecule force spectroscopy using polyproteins. *Chem. Soc. Rev.*, 41(14):4781–4796, 2012. doi:10.1039/c2cs35033e.
- [9] V. Barsegov, D. K. Klimov, and D. Thirumalai. Mapping the energy landscape of biomolecules using single molecule force correlation spectroscopy: theory and applications. *Biophys. J.*, 90(11):3827–3841, 2006. doi:10.1529/biophysj.105.075937.
- [10] I. Popa, P. Kosuri, J. Alegre-Cebollada, S. Garcia-Manyes, and J. M. Fernandez. Force dependency of biochemical reactions measured by single-molecule force-clamp spectroscopy. *Nat. Protoc.*, 8(7):1261–1276, 2013. doi:10.1038/nprot.2013.056.
- [11] X. Zhang, L. Ma, and Y. Zhang. High-resolution optical tweezers for single-molecule manipulation. *Yale J. Biol. Med.*, 86(3):367–383, 2013.
- [12] J. R. Moffitt, Y. R. Chemla, S. B. Smith, and C. Bustamante. Recent advances in optical tweezers. *Annu. Rev. Biochem.*, 77:205–228, 2008. doi:10.1146/annurev.biochem.77.043007.090225.
- [13] M. Capitanio and F. S. Pavone. Interrogating biology with force: single molecule high-resolution measurements with optical tweezers. *Biophys. J.*, 105(6):1293–1303, 2013. doi:10.1016/j.bpj.2013.08.007.
- [14] I. De Vlaminck and C. Dekker. Recent advances in magnetic tweezers. *Annu. Rev. Biophys.*, 41:453–472, 2012. doi:10.1146/annurev-biophys-122311-100544.
- [15] S. Le, R. Liu, C. T. Lim, and J. Yan. Uncovering mechanosensing mechanisms at the single protein level using magnetic tweezers. *Methods*, 94:13–18, 2015. doi:10.1016/j.ymeth.2015.08.020.
- [16] Devrim Kilinc and Gil U Lee. Advances in magnetic tweezers for single molecule and cell biophysics. *Integr. Qology Quant. Biosci. from Nano to Macro*, 6(1):27–34, 2014. doi:10.1039/c3ib40185e.
- [17] T. R. Strick, M.-N. Dessinges, G. Charvin, N. H. Dekker, J.-F. Allemand, D. Bensimon, and V. Croquette. Stretching of macromolecules and proteins. *Reports Prog. Phys.*, 66(1):1–45, 2003. doi:10.1088/0034-4885/66/1/201.

- [18] A. V. Finkelstein and O. B. Ptitsyn. *Protein Physics*. Elsevier, 2002. doi:10.1016/B978-012256781-0/50017-2.
- [19] M. Lieberman, A. D. Marks, C. M. Smith, and D. B. Marks. *Marks' Essential Medical Biochemistry*. Lippincott Williams & Wilkins, Baltimore, USA, 2007.
- [20] L. Pauling and R. B. Corey. Configuration of polypeptide chains. *Nature*, 168(4274):550–551, 1951. doi:10.1038/168550a0.
- [21] A. S. Edison. Linus Pauling and the planar peptide bond. *Nat. Struct. Biol.*, 8(3):201–202, 2001. doi:10.1038/84921.
- [22] R. Petrosyan, C. A. Bippes, S. Walheim, D. Harder, D. Fotiadis, T. Schimmel, D. Alsteens, and D. J. Müller. Single-molecule force spectroscopy of membrane proteins from membranes freely spanning across nanoscopic pores. *Nano Lett.*, 15(5):3624–3633, 2015. doi:10.1021/acs.nanolett.5b01223.
- [23] M. Pfreundschuh, D. Alsteens, R. Wieneke, C. Zhang, S. R. Coughlin, R. Tampé, B. K. Kobilka, and D. J. Müller. Identifying and quantifying two ligand-binding sites while imaging native human membrane receptors by AFM. *Nat. Commun.*, 6(8857):1–7, jan 2015. doi:10.1038/ncomms9857.
- [24] C. A. Bippes and D. J. Müller. High-resolution atomic force microscopy and spectroscopy of native membrane proteins. *Reports Prog. Phys.*, 74(8):086601(1)–086601(43), aug 2011. doi:10.1088/0034-4885/74/8/086601.
- [25] A. Kedrov, H. Janovjak, K. T. Sapra, and D. J. Müller. Deciphering molecular interactions of native membrane proteins by single-molecule force spectroscopy. *Annu. Rev. Biophys. Biomol. Struct.*, 36:233–260, 2007. doi:10.1146/annurev.biophys.36.040306.132640.
- [26] A. Engel and H. E. Gaub. Structure and mechanics of membrane proteins. *Annu. Rev. Biochem.*, 77:127–148, 2008. doi:10.1146/annurev.biochem.77.062706.154450.
- [27] A. M. Whited and P. S.-H. Park. Atomic force microscopy: a multifaceted tool to study membrane proteins and their interactions with ligands. *Biochim. Biophys. Acta*, 1838(1 Pt A):56–68, 2014. doi:10.1016/j.bbamem.2013.04.011.
- [28] Y. Shan and H. Wang. The structure and function of cell membranes examined by atomic force microscopy and single-molecule force spectroscopy. *Chem. Soc. Rev.*, 44(11):3617–3638, 2015. doi:10.1039/c4cs00508b.
- [29] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.*, 48:545–600, 1997. doi:10.1146/annurev.physchem.48.1.545.
- [30] M. Schlierf and M. Rief. Single-molecule unfolding force distributions reveal a funnel-shaped energy landscape. *Biophys. J.*, 90(4):L33–L35, 2006. doi:10.1529/biophysj.105.077982.
- [31] S. H. White and W. C. Wimley. Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.*, 28:319–365, 1999. doi:10.1146/annurev.biophys.28.1.319.
- [32] C. N. Pace, G. R. Grimsley, J. M. Scholtz, and K. L. Shaw. Protein stability. *Encycl. Life Sci.*, 2014. doi:DOI:10.1002/9780470015902.a0003002.pub3.
- [33] C. N. Pace. Conformational stability of globular proteins. *Trends Biochem. Sci.*, 15(1):14–17, 1990. doi:10.1016/0968-0004(90)90124-T.
- [34] C. N. Pace, B. A. Shirley, M. McNutt, and K. Gajiwala. Forces contributing to the conformational stability of proteins. *FASEB J.*, 10(1):75–83, 1996.
- [35] C. N. Pace, J. M. Scholtz, and G. R. Grimsley. Forces stabilizing proteins. *FEBS Lett.*, 588(14):2177–2184, 2014. doi:10.1016/j.febslet.2014.05.006.

- [36] L Pauling and R B Corey. The pleated sheet, a new layer configuration of polypeptide chains. *Proc. Natl. Acad. Sci. U. S. A.*, 37(5):251–256, 1951. doi:10.1073/pnas.37.5.251.
- [37] L. Pauling, R. B. Corey, and H. R. Branson. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.*, 37(4):205–211, 1951. doi:10.1073/pnas.37.4.205.
- [38] R. L. Baldwin. The new view of hydrophobic free energy. *FEBS Lett.*, 587(8):1062–1066, 2013. doi:10.1016/j.febslet.2013.01.006.
- [39] W Kauzmann. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.*, 14:1–63, 1959. doi:10.1016/S0065-3233(08)60608-7.
- [40] R. L. Baldwin. *Weak interactions in protein folding: hydrophobic free energy, van der Waals interactions, peptide hydrogen bonds, and peptide solvation*, volume 1. Wiley-VCH Verlag GmbH, 2008. doi:10.1002/9783527619498.ch6.
- [41] L. H. Kapcha and P. J. Rossky. A simple atomic-level hydrophobicity scale reveals protein interfacial structure. *J. Mol. Biol.*, 426(2):484–498, 2014. doi:10.1016/j.jmb.2013.09.039.
- [42] J. Kyte and R. F. Doolittle. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, 157(1):105–132, 1982.
- [43] S. C. Hue. Amino acid side-chain hydrophobicity. *Encycl. Life Sci.*, 2002. doi:10.1038/npg.els.0003005.
- [44] W. C. Wimley and S. H. White. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.*, 3(10):842–848, 1996. doi:10.1038/nsb1096-842.
- [45] L. Zhang, L. Wang, Y.-T. Kao, W. Qiu, Y. Yang, O. Okobiah, and D. Zhong. Mapping hydration dynamics around a protein surface. *Proc. Natl. Acad. Sci. U. S. A.*, 104(47):18461–18466, 2007. doi:10.1073/pnas.0707647104.
- [46] S. K. Pal and A. H. Zewail. Dynamics of water in biological recognition. *Chem. Rev.*, 104(4):2099–2123, 2004. doi:10.1021/cr0206891.
- [47] D. Zhong, S. K. Pal, and A. H. Zewail. Biological water: a critique. *Chem. Phys. Lett.*, 503(1-3):1–11, 2011. doi:10.1016/j.cplett.2010.12.077.
- [48] T. Yokomizo, M. Nakasako, T. Yamazaki, H. Shindo, and J. Higo. Hydrogen-bond patterns in the hydration structure of a protein. *Chem. Phys. Lett.*, 401(4-6):332–336, 2005. doi:10.1016/j.cplett.2004.11.071.
- [49] T. Yokomizo, J. Higo, and M. Nakasako. Patterns and networks of hydrogen-bonds in the hydration structure of human lysozyme. *Chem. Phys. Lett.*, 410(1-3):31–35, 2005. doi:10.1016/j.cplett.2005.04.072.
- [50] M. Chaplin. Do we underestimate the importance of water in cell biology? *Nat. Rev. Mol. Cell Biol.*, 7(11):861–866, 2006. doi:10.1038/nrm2021.
- [51] M. Koizumi, H. Hirai, T. Onai, K. Inoue, and M. Hirai. Collapse of the hydration shell of a protein prior to thermal unfolding. *J. Appl. Crystallogr.*, 40(s1):s175–s178, 2007. doi:10.1107/S0021889807003354.
- [52] I. Brovchenko, A. Krukau, N. Smolin, A. Oleinikova, A. Geiger, and R. Winter. Thermal breaking of spanning water networks in the hydration shell of proteins. *J. Chem. Phys.*, 123(22):224905(1)–224905(10), 2005. doi:10.1063/1.2121708.
- [53] N. Smolin, A. Oleinikova, I. Brovchenko, A. Geiger, and R. Winter. Properties of spanning water networks at protein surfaces. *J. Phys. Chem. B*, 109(21):10995–11005, 2005. doi:10.1021/jp050153e.

- [54] C. P. Moon and K. G. Fleming. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proc. Natl. Acad. Sci. U. S. A.*, 108(25):10174–10177, 2011. doi:10.1073/pnas.1103979108.
- [55] J. L. Fauchere and V. E. Pliska. Hydrophobic parameters II of amino acid side-chains from the partitioning of N-acetyl-amino acid amides. *Eur. J. Med. Chem.*, 18(4):369–375, 1983.
- [56] C. N. Pace. Contribution of the hydrophobic effect to globular protein stability. *J. Mol. Biol.*, 226(1):29–35, 1992. doi:10.1016/0022-2836(92)90121-Y.
- [57] M. Grandbois. How strong is a covalent bond? *Science*, 283(5408):1727–1730, 1999. doi:10.1126/science.283.5408.1727.
- [58] R. A. Maillard, G. Chistol, M. Sen, M. Righini, J. Tan, C. M. Kaiser, C. Hodges, A. Martin, and C. Bustamante. ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell*, 145(3):459–469, 2011. doi:10.1016/j.cell.2011.04.010.
- [59] A. O. Olivares, A. R. Nager, O. Iosefson, R. T. Sauer, and T. A. Baker. Mechanochemical basis of protein degradation by a double-ring AAA+ machine. *Nat. Struct. Mol. Biol.*, 21(10):871–875, 2014. doi:10.1038/nsmb.2885.
- [60] M.-E. Aubin-Tam, A. O. Olivares, R. T. Sauer, T. A. Baker, and M. J. Lang. Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell*, 145(2):257–267, 2011. doi:10.1016/j.cell.2011.03.036.
- [61] K. Svoboda, C. F. Schmidt, B. J. Schnapp, and S. M. Block. Direct observation of kinesin stepping by optical trapping interferometry. *Nature*, 365(6448):721–727, 1993. doi:10.1038/365721a0.
- [62] J. T. Finer, R. M. Simmons, and J. A. Spudich. Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature*, 368(6467):113–119, 1994. doi:10.1038/368113a0.
- [63] S. Chakrabarti, M. Hinczewski, and D. Thirumalai. Plasticity of hydrogen bond networks regulates mechanochemistry of cell adhesion complexes. *Proc. Natl. Acad. Sci. U. S. A.*, 111(25):9048–9053, 2014. doi:10.1073/pnas.1405384111.
- [64] E. M. Puchner, Al. Alexandrovich, A. L. Kho, U. Hensen, L. V. Schäfer, B. Brandmeier, F. Gräter, H. Grubmüller, H. E. Gaub, and M. Gautel. Mechanoenzymatics of titin kinase. *Proc. Natl. Acad. Sci. U. S. A.*, 105(36):13385–13390, 2008. doi:10.1073/pnas.0805034105.
- [65] E. C. Yusko and C. L. Asbury. Force is a signal that cells cannot ignore. *Mol. Biol. Cell*, 25(23):3717–3725, 2014. doi:10.1091/mbc.E13-12-0707.
- [66] V. Vogel. Mechanotransduction involving multimodular proteins: converting force into biochemical signals. *Annu. Rev. Biophys. Biomol. Struct.*, 35:459–88, 2006. doi:10.1146/annurev.biophys.35.040405.102013.
- [67] Y. Chen, S. E. Radford, and D. J. Brockwell. Force-induced remodelling of proteins and their complexes. *Curr. Opin. Struct. Biol.*, 30:89–99, 2015. doi:10.1016/j.sbi.2015.02.001.
- [68] N. A. Burnham, X. Chen, C. S. Hodges, G. A. Matei, E. J. Thoreson, C. J. Roberts, M. C. Davies, and S. J. B. Tendler. Comparison of calibration methods for atomic-force microscopy cantilevers. *Nanotechnology*, 14(1):1–6, 2003. doi:10.1088/0957-4484/14/1/301.
- [69] J. L. Hutter and J. Bechhoefer. Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.*, 64(7):1868–1873, 1993. doi:10.1063/1.1143970.
- [70] R. Levy and M. Maaloum. Measuring the spring constant of atomic force microscope cantilevers: thermal fluctuations and other methods. *Nanotechnology*, 13(1):33–37, 2002. doi:10.1088/0957-4484/13/1/307.

- [71] I. M. Malovichko. Measuring AFM cantilever stiffness from a thermal noise spectrum. *Bull. Russ. Acad. Sci. Phys.*, 77(8):972–974, 2013. doi:10.3103/S1062873813080248.
- [72] A. Labuda, J. R. Bates, and P. H. Grütter. The noise of coated cantilevers. *Nanotechnology*, 23(2):025503(1)–025503(9), 2012. doi:10.1088/0957-4484/23/2/025503.
- [73] A. Noy. *Handbook of Molecular Force Spectroscopy*. Springer US, New York, 2007. doi:10.1007/978-0-387-49989-5.
- [74] M. B. Viani, T. E. Sch^uffer, A. Chand, M. Rief, H. E. Gaub, and P. K. Hansma. Small cantilevers for force spectroscopy of single molecules. *J. Appl. Phys.*, 86(4):2258–2262, 1999. doi:10.1063/1.371039.
- [75] Z. Schumacher, Y. Miyahara, L. Aeschimann, and P. Grütter. Improved atomic force microscopy cantilever performance by partial reflective coating. *Beilstein J. Nanotechnol.*, 6(1):1450–146, 2015. doi:10.3762/bjnano.6.150.
- [76] M. S. Bull, R. M. A. Sullan, H. Li, and T. T. Perkins. Improved single molecule force spectroscopy using micromachined cantilevers. *ACS Nano*, 8(5):4984–4995, 2014. doi:10.1021/nn5010588.
- [77] R. M. A. Sullan, A. B. Churnside, D. M. Nguyen, M. S. Bull, and T. T. Perkins. Atomic force microscopy with sub-picoNewton force stability for biological applications. *Methods*, 60(2):131–141, 2013. doi:10.1016/j.ymeth.2013.03.029.
- [78] T. Hoffmann, K. M. Tych, D. J. Brockwell, and L. Dougan. Single-molecule force spectroscopy identifies a small cold shock protein as being mechanically robust. *J. Phys. Chem. B*, 117(6):1819–1826, 2013. doi:10.1021/jp310442s.
- [79] M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez. Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl. Acad. Sci. U. S. A.*, 96(7):3694–3699, 1999. doi:10.1073/pnas.96.7.3694.
- [80] T. Hoffmann, K. M. Tych, T. Crosskey, B. Schiffrin, D. J. Brockwell, and L. Dougan. Rapid and robust polyprotein production facilitates single-molecule mechanical characterization of β -barrel assembly machinery polypeptide transport associated domains. *ACS Nano*, 9(9):8811–8821, 2015. doi:10.1021/acsnano.5b01962.
- [81] P. Zheng and H. Li. Direct measurements of the mechanical stability of zinc-thiolate bonds in rubredoxin by single-molecule atomic force microscopy. *Biophys. J.*, 101(6):1467–1473, 2011. doi:10.1016/j.bpj.2011.08.021.
- [82] M. Carrion-Vazquez, A. F. Oberhauser, T. E. Fisher, P. E. Marszalek, H. B. Li, and J. M. Fernandez. Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Prog. Biophys. Mol. Biol.*, 74(1-2):63–91, 2000. doi:10.1016/S0079-6107(00)00017-1.
- [83] H. Dietz, M. Bertz, M. Schlierf, F. Berkemeier, T. Bornschlöggl, J. P. Junker, and M. Rief. Cysteine engineering of polyproteins for single-molecule force spectroscopy. *Nat. Protoc.*, 1(1):80–84, 2006. doi:10.1038/nprot.2006.12.
- [84] G. Yang, C. Cecconi, W. A. Baase, I. R. Vetter, W. A. Breyer, J. A. Haack, B. W. Matthews, F. W. Dahlquist, and C. Bustamante. Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme. *Proc. Natl. Acad. Sci. U. S. A.*, 97(1):139–144, 2000. doi:10.1073/pnas.97.1.139.
- [85] P. E. Marszalek and Y. F. Duf^rêne. Stretching single polysaccharides and proteins using atomic force microscopy. *Chem. Soc. Rev.*, 41(9):3523–3534, 2012. doi:10.1039/c2cs15329g.
- [86] R. L. Willett, K. W. Baldwin, K. W. West, and L. N. Pfeiffer. Differential adhesion of amino acids to inorganic surfaces. *Proc. Natl. Acad. Sci. U. S. A.*, 102(22):7817–7822, 2005. doi:10.1073/pnas.0408565102.

- [87] I. Popa, R. Berkovich, J. Alegre-Cebollada, C. L. Badilla, J. A. Rivas-Pardo, Y. Taniguchi, M. Kawakami, and J. M. Fernandez. Nanomechanics of HaloTag tethers. *J. Am. Chem. Soc.*, 135(34):12762–12771, 2013. doi:10.1021/ja4056382.
- [88] Y. Taniguchi and M. Kawakami. Application of HaloTag protein to covalent immobilization of recombinant proteins for single molecule force spectroscopy. *Langmuir*, 26(13):10433–10436, 2010. doi:10.1021/la101658a.
- [89] R. B. Best, B. Li, A. Steward, V. Daggett, and J. Clarke. Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophys. J.*, 81(4):2344–2356, 2001. doi:10.1016/s0006-3495(01)75881-x.
- [90] D. P. Sadler, E. Petrik, Y. Taniguchi, J. R. Pullen, M. Kawakami, S. E. Radford, and D. J. Brockwell. Identification of a mechanical rheostat in the hydrophobic core of protein L. *J. Mol. Biol.*, 393(1):237–248, 2009. doi:DOI10.1016/j.jmb.2009.08.015.
- [91] K. M. Tych, T. Hoffmann, J. Brockwell, and L. Dougan. Single molecule force spectroscopy reveals the temperature-dependent robustness and malleability of a hyperthermophilic protein. *Soft Matter*, 9:9016–9025, 2013. doi:10.1039/c3sm51439k.
- [92] D. J. Brockwell, G. S. Beddard, J. Clarkson, R. C. Zinober, A. W. Blake, J. Trinick, P. D. Olmsted, D. A. Smith, and S. E. Radford. The effect of core destabilization on the mechanical resistance of I27. *Biophys. J.*, 83(1):458–472, 2002. doi:10.1016/S0006-3495(02)75182-5.
- [93] E. M. Puchner, G. Franzen, M. Gautel, and H. E. Gaub. Comparing proteins by their unfolding pattern. *Biophys. J.*, 95(1):426–434, 2008. doi:10.1529/biophysj.108.129999.
- [94] T. Hugel, M. Rief, M. Seitz, H. E. Gaub, and R. R Netz. Highly stretched single polymers: atomic-force-microscope experiments versus ab-initio theory. *Phys. Rev. Lett.*, 94(4):048301(1)–048301(4), 2005. doi:10.1103/PhysRevLett.94.048301.
- [95] P. D. Bosshart, P. L. T. M. Frederix, and A. Engel. Reference-free alignment and sorting of single-molecule force spectroscopy data. *Biophys. J.*, 102(9):2202–2211, 2012. doi:10.1016/j.bpj.2012.03.027.
- [96] M. Fixman and J. Kovac. Polymer conformational statistics. III. Modified Gaussian models of stiff chains. *J. Chem. Phys.*, 58(4):1564–1568, 1973. doi:10.1063/1.1679396.
- [97] J. Mark. *Physical Properties of Polymers Handbook*. Cambridge University Press, Cambridge, 2004.
- [98] Harold A. Scheraga. *Protein Structure*. Academic Press, New York, 1961.
- [99] C. Storm and P. C. Nelson. Theory of high-force DNA stretching and overstretching. *Phys. Rev. E*, 67(5):051906(1)–051906(12), 2003. doi:http://dx.doi.org/10.1103/PhysRevE.67.051906.
- [100] F. Gräter, P. Heider, R. Zangi, and B. J. Berne. Dissecting entropic coiling and poor solvent effects in protein collapse. *J. Am. Chem. Soc.*, 130(35):11578–11579, 2008. doi:10.1021/ja802341q.
- [101] T. Kawakatsu. *Statistical Physics of Polymers: An Introduction*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
- [102] C. Bouchiat, M. D. Wang, J. Allemand, T. Strick, S. M. Block, and V. Croquette. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys. J.*, 76(1):409–413, 1999. doi:http://dx.doi.org/10.1016/S0006-3495(99)77207-3.
- [103] G. C. K. Roberts. *Encyclopedia of Biophysics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. doi:10.1007/978-3-642-16712-6.
- [104] J. F. Marko and E. D. Siggia. Stretching DNA. *Macromolecules*, 28(26):8759–8770, 1995. doi:10.1021/ma00130a008.

- [105] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith. Entropic elasticity of lambda-phage DNA. *Science*, 265(5178):1599–1600, 1994. doi:10.1126/science.8079175.
- [106] P. E. Marszalek, H. Lu, H. B. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J M Fernandez. Mechanical unfolding intermediates in titin modules. *Nature*, 402(6757):100–103, 1999. doi:10.1038/47083.
- [107] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276(5315):1109–1112, 1997. doi:10.1126/science.276.5315.1109.
- [108] T. Su and P. K. Purohit. Mechanics of forced unfolding of proteins. *Acta Biomater.*, 5(6):1855–1863, 2009. doi:10.1016/j.actbio.2009.01.038.
- [109] S. B. Smith, Y. Cui, and C. Bustamante. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, 271(5250):795–799, 1996. doi:10.1126/science.271.5250.795.
- [110] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block. Stretching DNA with optical tweezers. *Biophys. J.*, 72(3):1335–1346, 1997. doi:10.1016/S0006-3495(97)78780-0.
- [111] N. M. Toan, D. Marenduzzo, and C. Micheletti. Inferring the diameter of a biopolymer from its stretching response. *Biophys. J.*, 89(1):80–86, 2005. doi:10.1529/biophysj.104.058081.
- [112] L. Livadaru, R. R. Netz, and H. J. Kreuzer. Stretching response of discrete semiflexible polymers. *Macromolecules*, 36(10):3732–3744, may 2003. doi:10.1021/ma020751g.
- [113] D. W. Urry, T. Hugel, M. Seitz, H. E. Gaub, L. Sheiba, J. Dea, J. Xu, and T. Parker. Elastin: a representative ideal protein elastomer. *Philos. Trans. R. Soc. B Biol. Sci.*, 357(1418):169–184, 2002. doi:10.1098/rstb.2001.1023.
- [114] N. M. Toan and D. Thirumalai. On the origin of the unusual behavior in the stretching of single-stranded DNA. *J. Chem. Phys.*, 136(23):235103(1)–235103(5), 2012. arXiv:arXiv:1205.6650v2, doi:10.1063/1.4729371.
- [115] R. Liu, M. Roman, and G. Yang. Correction of the viscous drag induced errors in macromolecular manipulation experiments using atomic force microscope. *Rev. Sci. Instrum.*, 81(6):063703(1)–063703(5), 2010. doi:10.1063/1.3436646.
- [116] H. Janovjak, D. J. Muller, and A. D. L. Humphris. Molecular force modulation spectroscopy revealing the dynamic response of single bacteriorhodopsins. *Biophys. J.*, 88(2):1423–1431, 2005. doi:10.1529/biophysj.104.052746.
- [117] J.-M. Yuan, C.-L. Chyan, H.-X. Zhou, T.-Y. Chung, H. Peng, G. Ping, and G. Yang. The effects of macromolecular crowding on the mechanical stability of protein molecules. *Protein Sci.*, 17(12):2156–2166, 2008. doi:10.1110/ps.037325.108.
- [118] J. Alcaraz, L. Buscemi, M. Puig-de Morales, J. Colchero, A. Baro, and D. Navajas. Correction of microrheological measurements of soft samples with atomic force microscopy for the hydrodynamic drag on the cantilever. *Langmuir*, 18(3):716–721, 2002. doi:10.1021/la0110850.
- [119] F. Rico, L. Gonzalez, I. Casuso, M. Puig-Vidal, and S. Scheuring. High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations. *Science*, 342(6159):741–743, 2013. doi:10.1126/science.1239764.
- [120] Z. T. Yew, M. Schlierf, M. Rief, and E. Paci. Direct evidence of the multidimensionality of the free-energy landscapes of proteins revealed by mechanical probes. *Phys. Rev. E*, 81(3):031923(1)–031923(4), 2010. doi:10.1103/PhysRevE.81.031923.

- [121] C. Hyeon, M. Hinczewski, and D. Thirumalai. Evidence of disorder in biological molecules from single molecule pulling experiments. *Phys. Rev. Lett.*, 112(13):138101(1)–138101(5), 2014. doi:10.1103/PhysRevLett.112.138101.
- [122] B. Khatri, Z. T. Yew, S. Krivov, T. McLeish, and E. Paci. Fluctuation power spectra reveal dynamical heterogeneity of peptides. *J. Chem. Phys.*, 133(1):015101(1)–015101(6), 2010. doi:10.1063/1.3456552.
- [123] G. I. Bell. Models for specific adhesion of cells to cells. *Science*, 200(4342):618–627, 1978. doi:10.1126/science.347575.
- [124] O. K. Dudko, G. Hummer, and A. Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. U. S. A.*, 105(41):15755–15760, 2008. doi:10.1073/pnas.0806085105.
- [125] R. W. Friddle, A. Noy, and J. J. De Yoreo. Interpreting the widespread nonlinear force spectra of intermolecular bonds. *Proc. Natl. Acad. Sci. U. S. A.*, 109(34):13573–13578, 2012. doi:10.1073/pnas.1202946109.
- [126] Z. T. Yew, P. D. Olmsted, and E. Paci. Free energy landscapes of proteins: insights from mechanical probes. In *Single-Molecule Biophys.*, pages 395–417. John Wiley & Sons, Inc., 2011. doi:10.1002/9781118131374.ch14.
- [127] F. T. Hane, S. J. Attwood, and Z. Leonenko. Comparison of three competing dynamic force spectroscopy models to study binding forces of amyloid-beta (1-42). *Soft Matter*, 10(12):1924–1930, 2014. doi:10.1039/c3sm52257a.
- [128] S. N. Zhurkov. Kinetic concept of strength of solids. *Int. J. Fract. Mech.*, 1(4):311–322, 1965. doi:10.1007/BF00962961.
- [129] E. Evans and K. Ritchie. Dynamic strength of molecular adhesion bonds. *Biophys. J.*, 72(4):1541–1555, 1997. doi:10.1016/S0006-3495(97)78802-7.
- [130] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940. doi:10.1016/S0031-8914(40)90098-2.
- [131] E. Evans. Probing the relation between force–lifetime–and chemistry in single molecular bonds. *Annu. Rev. Biophysics Biomolecular Struct.*, 30:105–128, 2001. doi:10.1146/annurev.biophys.30.1.105.
- [132] D. J. Brockwell, G. S. Beddard, E. Paci, D. K. West, P. D. Olmsted, D. A. Smith, and S. E. Radford. Mechanically unfolding the small, topologically simple protein L. *Biophys. J.*, 89(1):506–519, 2005. doi:10.1529/biophysj.105.061465.
- [133] O. K. Dudko, G. Hummer, and A. Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.*, 96(10):1–4, 2006. doi:10.1103/PhysRevLett.96.108101.
- [134] G. Hummer and A. Szabo. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophys. J.*, 85(1):5–15, 2003. doi:10.1016/S0006-3495(03)74449-X.
- [135] Y. J. Zhang and O. K. Dudko. A transformation for the mechanical fingerprints of complex biomolecular interactions. *Proc. Natl. Acad. Sci. U. S. A.*, 110(41):16432–16437, 2013. doi:DOI10.1073/pnas.1309101110.
- [136] G. Hummer and A. Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. U. S. A.*, 98(7):3658–3661, 2001. doi:10.1073/pnas.071034098.
- [137] A. Imparato, F. Sbrana, and M. Vassalli. Reconstructing the free-energy landscape of a polyprotein by single-molecule experiments. *Europhys. Lett.*, 82(58006):1–5, 2008. doi:10.1209/0295-5075/82/58006.

- [138] G. Hummer and A. Szabo. Free energy profiles from single-molecule pulling experiments. *Proc. Natl. Acad. Sci. U. S. A.*, 107(50):21441–21446, dec 2010. doi:10.1073/pnas.1015661107.
- [139] J. C. M. Gebhardt, T. Bornschlöggl, and M. Rief. Full distance-resolved folding energy landscape of one single protein molecule. *Proc. Natl. Acad. Sci. U. S. A.*, 107(5):2013–2008, 2010. doi:10.1073/pnas.0909854107.
- [140] M. Kessler, K. E. Gottschalk, H. Janovjak, D. J. Muller, and H. E. Gaub. Bacteriorhodopsin folds into the membrane against an external force. *J. Mol. Biol.*, 357(2):644–654, 2006. doi:10.1016/j.jmb.2005.12.065.
- [141] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E*, 56(5):5018–5035, 1997. doi:http://dx.doi.org/10.1103/PhysRevE.56.5018.
- [142] G. E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, 60(3):2721–2726, 1999. arXiv:9901352, doi:10.1103/PhysRevE.60.2721.
- [143] G. Hummer and A. Szabo. Free energy surfaces from single-molecule force spectroscopy. *Acc. Chem. Res.*, 38(7):504–513, 2005. doi:10.1021/ar040148d.
- [144] O. Braun, A. Hanke, and U. Seifert. Probing molecular free energy landscapes by periodic loading. *Phys. Rev. Lett.*, 93(15):158105(1)–158105(4), 2004. doi:10.1103/PhysRevLett.93.158105.
- [145] A. Imparato and L. Peliti. Evaluation of free energy landscapes from manipulation experiments. *J. Stat. Mech. Theory Exp.*, 2006(03):P03005–P03026, 2006. doi:10.1088/1742-5468/2006/03/P03005.
- [146] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski’s equality. *Science*, 296(5574):1832–1835, 2002. doi:10.1126/science.1071152.
- [147] A. N. Gupta, A. Vincent, K. Neupane, H. Yu, F. Wang, and M.T. Woodside. Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements. *Nat. Phys.*, 7(8):631–634, 2011. doi:10.1038/nphys2022.
- [148] D. K. West, P. D. Olmsted, and E. Paci. Free energy for protein folding from nonequilibrium simulations using the Jarzynski equality. *J. Chem. Phys.*, 125(20):204910(1)–204910(7), 2006. doi:10.1063/1.2393232.
- [149] S. Luccioli, A. Imparato, and A. Torcini. Free-energy landscape of mechanically unfolded model proteins: extended Jarzynski versus inherent structure reconstruction. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.*, 78(3 Pt 1):031907(1)–031907(15), 2008. doi:http://dx.doi.org/10.1103/PhysRevE.78.031907.
- [150] S. Mitternacht, S. Luccioli, A. Torcini, A. Imparato, and A. Irbäck. Changing the mechanical unfolding pathway of FnIII10 by tuning the pulling strength. *Biophys. J.*, 96(2):429–441, 2009. doi:10.1016/j.bpj.2008.09.043.
- [151] N. Harris, Y. Song, and C.-H. Kiang. Experimental free energy surface reconstruction from single-molecule force spectroscopy using Jarzynski’s equality. *Phys. Rev. Lett.*, 99(6):068101(1)–068101(4), 2007. doi:10.1103/PhysRevLett.99.068101.
- [152] J. Gore, F. Ritort, and C. Bustamante. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc. Natl. Acad. Sci. U. S. A.*, 100(22):12564–12569, 2003. doi:10.1073/pnas.1635159100.
- [153] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature*, 437(7056):231–234, 2005.

- [154] C. He, C. Hu, X. Hu, X. Hu, A. Xiao, T. T. Perkins, and H. Li. Direct observation of the reversible two-state unfolding and refolding of an α/β protein by single-molecule atomic force microscopy. *Angew. Chem. Int. Ed. Engl.*, 54(34):9921–9925, 2015. doi:10.1002/anie.201502938.
- [155] R. C. Zinober, D. J. Brockwell, G. S. Beddard, A. W. Blake, P. D. Olmsted, S. E. Radford, and D. A. Smith. Mechanically unfolding proteins: the effect of unfolding history and the supramolecular scaffold. *Protein Sci.*, 11(12):2759–2765, 2002. doi:10.1110/ps.0224602.
- [156] K. M. Tych, M. L. Hughes, J. Bourke, Y. Taniguchi, M. Kawakami, D. J. Brockwell, and L. Dougan. Optimizing the calculation of energy landscape parameters from single-molecule protein unfolding experiments. *Phys. Rev. E*, 91(1):012710(1)–012710(9), 2015. doi:10.1103/PhysRevE.91.012710.
- [157] M. Wolny, M. Batchelor, P. J. Knight, E. Paci, L. Dougan, and M. Peckham. Stable single α -helices are constant force springs in proteins. *J. Biol. Chem.*, 289(40):27825–27835, 2014. doi:10.1074/jbc.M114.585679.
- [158] M. S. Li. Secondary structure, mechanical stability, and location of transition state of proteins. *Biophys. J.*, 93(8):2644–2654, 2007. doi:10.1529/biophysj.107.106138.
- [159] H. B. Li and J. M. Fernandez. Mechanical design of the first proximal Ig domain of human cardiac titin revealed by single molecule force spectroscopy. *J. Mol. Biol.*, 334(1):75–86, 2003. doi:10.1016/j.jmb.2003.09.036.
- [160] A F Oberhauser, C Badilla-Fernandez, M Carrion-Vazquez, and J M Fernandez. The mechanical hierarchies of fibronectin observed with single-molecule AFM. *J. Mol. Biol.*, 319(2):433–447, 2002. doi:10.1016/S0022-2836(02)00306-6.
- [161] K. L. Fuson, L. Ma, R. B. Sutton, and A. F. Oberhauser. The c2 domains of human synaptotagmin 1 have distinct mechanical properties. *Biophys. J.*, 96(3):1083–1090, 2009. doi:10.1016/j.bpj.2008.10.025.
- [162] D. J. Brockwell, E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, and S. E. Radford. Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nat. Struct. Biol.*, 10(9):731–737, 2003. doi:10.1038/nsb968.
- [163] Y. Cao and H. B. Li. Polyprotein of GB1 is an ideal artificial elastomeric protein. *Nat. Mater.*, 6(2):109–114, 2007. doi:10.1038/Nmat1825.
- [164] D. Sharma, G. Feng, D. Khor, G. Z. Genchev, H. Lu, and H. B. Li. Stabilization provided by neighboring strands is critical for the mechanical stability of proteins. *Biophys. J.*, 95(8):3935–3942, 2008. doi:10.1529/biophysj.108.134072.
- [165] M. Carrion-Vazquez, H. B. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez. The mechanical stability of ubiquitin is linkage dependent. *Nat. Struct. Biol.*, 10(9):738–743, 2003. doi:10.1038/nsb965.
- [166] G. Arad-Haase, S. G. Chuartzman, S. Dagan, R. Nevo, M. Kouza, B. K. Mai, H. T. Nguyen, M. S. Li, and Z. Reich. Mechanical unfolding of acylphosphatase studied by single-molecule force spectroscopy and MD simulations. *Biophys. J.*, 99(1):238–247, 2010. doi:10.1016/j.bpj.2010.04.004.
- [167] M. Rief, J. Pascual, M. Saraste, and H. E. Gaub. Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *J. Mol. Biol.*, 286(2):553–561, 1999. doi:10.1006/jmbi.1998.2466.
- [168] S. P. Ng, R. W. S. Rounsevell, A. Steward, C. D. Geierhaas, P. M. Williams, E. Paci, and J. Clarke. Mechanical unfolding of TNfn3: the unfolding pathway of a fnIII domain probed by protein engineering, AFM and MD simulation. *J. Mol. Biol.*, 350(4):776–789, 2005. doi:10.1016/j.jmb.2005.04.070.
- [169] D. J. Brockwell. Probing the mechanical stability of proteins using the atomic force microscope. *Biochem. Soc. Trans.*, 35(6):1564–1568, 2007. doi:10.1042/BST0351564.

- [170] H. Lu and K. Schulten. Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins Struct. Funct. Bioinforma.*, 35(4):453–463, 1999. doi:10.1002/(SICI)1097-0134(19990601)35:4<453::AID-PROT9>3.0.CO;2-M.
- [171] D. K. Klimov and D. Thirumalai. Native topology determines force-induced unfolding pathways in globular proteins. *Proc. Natl. Acad. Sci. U. S. A.*, 97(13):7254–7259, 2000. doi:10.1073/pnas.97.13.7254.
- [172] H. Dietz, F. Berkemeier, M. Bertz, and M. Rief. Anisotropic deformation response of single protein molecules. *Proc. Natl. Acad. Sci. U. S. A.*, 103(34):12724–12728, 2006. doi:10.1073/pnas.0602995103.
- [173] H. Dietz and M. Rief. Elastic bond network model for protein unfolding mechanics. *Phys. Rev. Lett.*, 100(9):098101(1)–098101(4), mar 2008. doi:10.1103/PhysRevLett.100.098101.
- [174] E. A. Shank, C. Cecconi, J. W. Dill, S. Marqusee, and C. Bustamante. The folding cooperativity of a protein is controlled by its chain topology. *Nature*, 465(7298):637–40, 2010. doi:10.1038/nature09021.
- [175] B. Jagannathan, P. J. Elms, C. Bustamante, and S. Marqusee. Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein. *Proc. Natl. Acad. Sci. U. S. A.*, 109(44):17820–17825, 2012. doi:10.1073/pnas.1201800109.
- [176] Morten Bertz, Andrea Kunfermann, and Matthias Rief. Navigating the Folding Energy Landscape of Green Fluorescent Protein. *Angew. Chemie Int. Ed.*, 47(43):8192–8195, 2008. doi:10.1002/anie.200802987.
- [177] M. Bertz and M. Rief. Mechanical unfoldons as building blocks of maltose-binding protein. *J. Mol. Biol.*, 378(2):447–458, 2008. doi:10.1016/j.jmb.2008.02.025.
- [178] P.-C. Li and D. E. Makarov. Simulation of the mechanical unfolding of ubiquitin: Probing different unfolding reaction coordinates by changing the pulling geometry. *J. Chem. Phys.*, 121(10):4826–4832, 2004. doi:10.1063/1.1778152.
- [179] W. Zheng and P. Glenn. Probing the folded state and mechanical unfolding pathways of T4 lysozyme using all-atom and coarse-grained molecular simulation. *J. Chem. Phys.*, 142(3):035101(1)–035101(14), 2015. doi:10.1063/1.4905606.
- [180] Y. D. Li, G. Lamour, J. Gsponer, P. Zheng, and H. Li. The molecular mechanism underlying mechanical anisotropy of the protein GB1. *Biophys. J.*, 103(11):2361–2368, 2012. doi:10.1016/j.bpj.2012.10.035.
- [181] G. Radou, M. Enciso, S. Krivov, and E. Paci. Modulation of a protein free-energy landscape by circular permutation. *J. Phys. Chem. B*, 117(44):13743–13747, 2013. doi:10.1021/jp406818t.
- [182] Z. Zhuang, A. I. Jewett, P. Soto, and J.-E. Shea. The effect of surface tethering on the folding of the src-SH3 protein domain. *Phys. Biol.*, 6(1):015004(1)–015004(10), 2009. doi:10.1088/1478-3975/6/1/015004.
- [183] A. J. Wilcox, J. Choy, C. Bustamante, and A. Matouschek. Effect of protein structure on mitochondrial import. *Proc. Natl. Acad. Sci. U. S. A.*, 102(43):15435–15440, 2005. doi:10.1073/pnas.0507324102.
- [184] Q. Peng and H. Li. Atomic force microscopy reveals parallel mechanical unfolding pathways of T4 lysozyme: evidence for a kinetic partitioning mechanism. *Proc. Natl. Acad. Sci. U. S. A.*, 105(6):1885–1890, 2008. doi:10.1073/pnas.0706775105.
- [185] Ulf Hensen and Daniel J Müller. Mechanistic explanation of different unfolding behaviors observed for transmembrane and soluble β -barrel proteins. *Structure*, 21(8):1317–1324, 2013. doi:10.1016/j.str.2013.06.001.

- [186] T. Serdiuk, J. Sugihara, S. A. Mari, H. R. Kaback, and D. J. Müller. Observing a lipid-dependent alteration in single lactose permeases. *Structure*, 23(4):754–761, 2015. doi:10.1016/j.str.2015.02.009.
- [187] M. Zocher, J. J. Fung, B. K. Kobilka, and D. J. Müller. Ligand-specific interactions modulate kinetic, energetic, and mechanical properties of the human β 2 adrenergic receptor. *Structure*, 20(8):1391–1402, 2012. doi:10.1016/j.str.2012.05.010.
- [188] 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, and M. A. Nash. Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.*, 5(5635):1–8, 2014. doi:10.1038/ncomms6635.
- [189] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.*, 75(2):662–671, 1998. doi:10.1016/S0006-3495(98)77556-3.
- [190] H. Lu and K. Schulten. The key event in force-induced unfolding of titin’s immunoglobulin domains. *Biophys. J.*, 79(1):51–65, 2000. doi:10.1016/S0006-3495(00)76273-4.
- [191] M. Sikora, M. Cieplak, and J. I. Sulkowska. Bio-molecule stretching database, 2010.
- [192] M. Sikora, J. I. Sulkowska, B. S. Witkowski, and M. Cieplak. BSDB: the biomolecule stretching database. *Nucleic Acids Res.*, 39:D443–D450, 2011. doi:10.1093/nar/gkq851.
- [193] M. Sikora, J. I. Sulkowska, and M. Cieplak. Mechanical strength of 17,134 model proteins and cysteine slipknots. *PLoS Comput. Biol.*, 5(10):1–15, 2009. doi:10.1371/journal.pcbi.1000547.
- [194] H. Li, M. Carrion-vazquez, A. F. Oberhauser, P. E. Marszalek, and J. M. Fernandez. Point mutations alter the mechanical stability of immunoglobulin modules. *Nat. Struct. Biol.*, 7(12):1117–1120, 2000. doi:10.1038/81964.
- [195] T. Bu, H.-C. E. Wang, and H. Li. Single molecule force spectroscopy reveals critical roles of hydrophobic core packing in determining the mechanical stability of protein GB1. *Langmuir*, 28(33):12319–12325, 2012. doi:10.1021/la301940g.
- [196] M. Sikora and M. Cieplak. Cystine plug and other novel mechanisms of large mechanical stability in dimeric proteins. *Phys. Rev. Lett.*, 109(20):208101(1)–208101(5), 2012. doi:10.1103/PhysRevLett.109.208101.
- [197] A. M. Gabovich and M. S. Li. Mechanical stability of proteins. *J. Chem. Phys.*, 131(2):024121(1)–024121(5), 2009. doi:10.1063/1.3170940.
- [198] LLC Schrodinger. The PyMOL Molecular Graphics System, Version 1.7.4. aug 2010.
- [199] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 112(3):535–542, 1977. doi:10.1111/j.1432-1033.1977.tb11885.x.
- [200] S. Improtà, A. S. Politou, and A. Pastore. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure*, 4(3):323–337, 1996. doi:10.1016/S0969-2126(96)00036-6.
- [201] J. W. O’Neill, D. E. Kim, D. Baker, and K. Y. Zhang. Structures of the B1 domain of protein L from *Peptostreptococcus magnus* with a tyrosine to tryptophan substitution. *Acta Crystallogr.*, 57(4):480–487, 2001. doi:10.1107/S0907444901000373.
- [202] L. J. Shimon, E. A. Bayer, E. Morag, R. Lamed, S. Yaron, Y. Shoham, and F. Frolow. A cohesin domain from *Clostridium thermocellum*: the crystal structure provides new insights into cellulosome assembly. *Structure*, 5(3):381–390, 1997. doi:10.1016/S0969-2126(97)00195-0.

- [203] W. Kremer, B. Schuler, S. Harrieder, M. Geyer, W. Gronwald, C. Welker, R. Jaenicke, and H. R. Kalbitzer. Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritima*. *Eur. J. Biochem.*, 268(9):2527–2539, 2001. doi:10.1046/j.1432-1327.2001.02127.x.
- [204] J. Pascual, M. Pfuhl, D. Walther, M. Saraste, and M. Nilges. Solution structure of the spectrin repeat: a left-handed antiparallel triple-helical coiled-coil. *J. Mol. Biol.*, 273(3):740–751, 1997. doi:10.1006/jmbi.1997.1344.
- [205] M. Bycroft, S. Ludvigsen, A. R. Fersht, and F. M. Poulsen. Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy. *Biochemistry*, 30(35):8697–8701, 1991. doi:10.1021/bi00099a030.
- [206] S. B. Fowler, R. B. Best, J. L. Toca Herrera, T. J. Rutherford, A. Steward, E. Paci, M. Karplus, J. Clarke, and J. L. T. Herrera. Mechanical unfolding of a titin Ig domain: structure of unfolding intermediate revealed by combining AFM, molecular dynamics simulations, NMR and protein engineering. *J. Mol. Biol.*, 322(4):841–849, 2002. doi:10.1016/S0022-2836(02)00805-7.
- [207] L. W. Li, H. H.-L.L Huang, C. L. Badilla, and J. M. Fernandez. Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *J. Mol. Biol.*, 345(4):817–826, 2005. doi:10.1016/j.jmb.2004.11.021.
- [208] Robert B. Best, Susan B. Fowler, José L. Toca Herrera, Annette Steward, Emanuele Paci, Jane Clarke, José L Toca Herrera, Annette Steward, Emanuele Paci, and Jane 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):867–877, 2003. doi:10.1016/S0022-2836(03)00618-1.
- [209] P. W. Fenimore, H. Frauenfelder, B. H. McMahon, and F. G. Parak. Slaving: solvent fluctuations dominate protein dynamics and functions. *Proc. Natl. Acad. Sci. U. S. A.*, 99(25):16047–16051, 2002. doi:10.1073/pnas.212637899.
- [210] G. Caliskan, D. Mechtani, J. H. Roh, A. Kisliuk, A. P. Sokolov, S. Azzam, M. T. Cicerone, S. Lin-Gibson, and I. Peral. Protein and solvent dynamics: how strongly are they coupled? *J. Chem. Phys.*, 121(4):1978–1983, 2004. doi:10.1063/1.1764491.
- [211] P. Cioni and G. B. Strambini. Effect of heavy water on protein flexibility. *Biophys. J.*, 82(6):3246–3253, 2002. doi:10.1016/S0006-3495(02)75666-X.
- [212] D. L. Pincus and D. Thirumalai. Crowding effects on the mechanical stability and unfolding pathways of ubiquitin. *J. Phys. Chem. B*, 113(1):359–368, 2009. arXiv:0811.0781, doi:10.1021/jp807755b.
- [213] K. Tych and L. Dougan. A single-molecule approach to explore the role of the solvent environment in protein folding. In *Proteins Solut. Interfaces*, pages 315–334. John Wiley & Sons, Inc., 2013. doi:10.1002/9781118523063.ch15.
- [214] S. Garcia-Manyes, L. Dougan, and J. M. Fernandez. Osmolyte-induced separation of the mechanical folding phases of ubiquitin. *Proc. Natl. Acad. Sci. U. S. A.*, 106(26):10540–10545, 2009. doi:10.1073/pnas.0902090106.
- [215] Y. Cao and H. Li. How do chemical denaturants affect the mechanical folding and unfolding of proteins? *J. Mol. Biol.*, 375(1):316–324, 2008. doi:10.1016/j.jmb.2007.10.024.
- [216] D. Aioanei, M. Bruciale, I. Tessari, L. Bubacco, and B. Samorì. Worm-like Ising model for protein mechanical unfolding under the effect of osmolytes. *Biophys. J.*, 102(2):342–350, 2012. doi:10.1016/j.bpj.2011.12.007.
- [217] L. Ma, M. Xu, and A. F. Oberhauser. Naturally occurring osmolytes modulate the nano-mechanical properties of polycystic kidney disease (PKD) domain. *J. Biol. Chem.*, 285(49):38438–38443, 2010. doi:10.1074/jbc.M110.183913.

- [218] E. P. O'Brien, B. R. Brooks, and D. Thirumalai. Effects of pH on proteins: predictions for ensemble and single-molecule pulling experiments. *J. Am. Chem. Soc.*, 134(2):979–987, 2012. doi:10.1021/ja206557y.
- [219] P. Zheng, Y. Cao, T. J. Bu, S. K. K. Straus, and H. B. Li. Single molecule force spectroscopy reveals that electrostatic interactions affect the mechanical stability of proteins. *Biophys. J.*, 100(6):1534–1541, 2011. doi:DOI10.1016/j.bpj.2011.01.062.
- [220] B. Cheng, S. Wu, S. Liu, P. Rodriguez-Aliaga, J. Yu, and S. Cui. Protein denaturation at a single-molecule level: the effect of nonpolar environments and its implications on the unfolding mechanism by proteases. *Nanoscale*, 7(7):2970–2977, 2015. doi:10.1039/c4nr07140a.
- [221] M. Schlierf and M. Rief. Temperature softening of a protein in single-molecule experiments. *J. Mol. Biol.*, 354(2):497–503, 2005. doi:10.1016/j.jmb.2005.09.070.
- [222] C. Hyeon and D. Thirumalai. Multiple barriers in forced rupture of protein complexes. *J. Chem. Phys.*, 137(5):055103(1)–055103(8), 2012. arXiv:arXiv:1204.1418v1, doi:10.1063/1.4739747.
- [223] C. Hyeon and D. Thirumalai. Measuring the energy landscape roughness and the transition state location of biomolecules using single molecule mechanical unfolding experiments. *J. Phys. Condens. Matter*, 19(11):113101(1)–113101(27), 2006. arXiv:0612433, doi:10.1088/0953-8984/19/11/113101.
- [224] L. J. Lapidus, W. A. Eaton, and J. Hofrichter. Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci. U. S. A.*, 97(13):7220–7225, 2000. doi:10.1073/pnas.97.13.7220.
- [225] C. B Hyeon and D. Thirumalai. Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments? *Proc. Natl. Acad. Sci. U. S. A.*, 100(18):10249–10253, 2003. doi:10.1073/pnas.1833310100.
- [226] Y. Taniguchi, D. J. Brockwell, and M. Kawakami. The effect of temperature on mechanical resistance of the native and intermediate states of I27. *Biophys. J.*, 95(11):5296–5305, 2008. doi:10.1529/biophysj.108.141275.
- [227] R. B. Best, S. B. Fowler, J. L. Toca-Herrera, and J. Clarke. A simple method for probing the mechanical unfolding pathway of proteins in detail. *Proc. Natl. Acad. Sci. U. S. A.*, 99(19):12143–12148, 2002. doi:10.1073/pnas.192351899.
- [228] A. R. Fersht, A. Matouschek, and L. Serrano. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.*, 224(3):771–782, 1992.
- [229] T. Hoffmann, K. M. Tych, M. L. Hughes, D. J. Brockwell, and L. Dougan. Towards design principles for determining the mechanical stability of proteins. *Phys. Chem. Chem. Phys.*, 15(38):15767–15780, 2013. doi:10.1039/C3CP52142G.
- [230] G. Yoon, H.-J. Park, S. Na, and K. Eom. Mesoscopic model for mechanical characterization of biological protein materials. *J. Comput. Chem.*, 30(6):873–880, 2009. doi:10.1002/jcc.21107.
- [231] M. Guthold, W. Liu, E. A. Sparks, L. M. Jawerth, L. Peng, M. Falvo, R. Superfine, R. R. Hantgan, and S. T. Lord. A comparison of the mechanical and structural properties of fibrin fibers with other protein fibers. *Cell Biochem. Biophys.*, 49(3):165–181, 2007. doi:10.1007/s12013-007-9001-4.
- [232] A. Ikai. Local rigidity of a protein molecule. *Biophys. Chem.*, 116(3):187–191, 2005. doi:10.1016/j.bpc.2005.04.003.
- [233] S. Keten and M. J. Buehler. Geometric confinement governs the rupture strength of H-bond assemblies at a critical length scale. *Nano Lett.*, 8(2):743–748, 2008. doi:10.1021/nl0731670.

- [234] A. M. R. D. Graff, G. Shannon, D.W. Farrell, P. M. Williams, and M. F. Thorpe. Protein unfolding under force : crack propagation in a network. *Biophys. J.*, 101(3):736–744, 2011. doi:10.1016/j.bpj.2011.05.072.
- [235] Sinan Keten and Markus J. Buehler. Strength limit of entropic elasticity in beta-sheet protein domains. *Phys. Rev. E*, 78(6):061913(1)–061913(7), 2008. doi:10.1103/PhysRevE.78.061913.
- [236] S. Keten and M. J. Buehler. Asymptotic strength limit of hydrogen-bond assemblies in proteins at vanishing pulling rates. *Phys. Rev. Lett.*, 100(19):198301(1)–198301(4), 2008. doi:10.1103/PhysRevLett.100.198301.
- [237] E. G. Baker, G. J. Bartlett, M. P. Crump, R. B. Sessions, N. Linden, C. F. J. Faul, and D. N. Woolfson. Local and macroscopic electrostatic interactions in single α -helices. *Nat. Chem. Biol.*, 11(3):221–228, 2015. doi:10.1038/nchembio.1739.
- [238] K. M. Tych, M. Batchelor, T. Hoffmann, M. C. Wilson, E. Paci, D. J. Brockwell, and L. Dougan. Tuning protein mechanics through an ionic cluster graft from an extremophilic protein. *Soft Matter*, 12:2688–2699, 2016. doi:10.1039/C5SM02938D.
- [239] K. M. Tych, T. Hoffmann, M. Batchelor, M. L. Hughes, K. E. Kendrick, D. L. Walsh, M. Wilson, D. J. Brockwell, and L. Dougan. Life in extreme environments: single molecule force spectroscopy as a tool to explore proteins from extremophilic organisms. *Biochem. Soc. Trans.*, 43(2):179–185, 2015. doi:10.1042/BST20140274.
- [240] N. Kong and H. Li. Protein fragment reconstitution as a driving force for self-assembling reversible protein hydrogels. *Adv. Funct. Mater.*, 25(35):5593–5601, 2015. doi:10.1002/adfm.201502277.
- [241] N. Kong, Q. Peng, and H. Li. Rationally designed dynamic protein hydrogels with reversibly tunable mechanical properties. *Adv. Funct. Mater.*, 24(46):7310–7317, 2014. doi:10.1002/adfm.201402205.
- [242] J. Fang, A. Mehlich, N. Koga, J. Huang, R. Koga, X. Gao, C. Hu, C. Jin, M. Rief, J. Kast, D. Baker, and H. Li. Forced protein unfolding leads to highly elastic and tough protein hydrogels. *Nat. Commun.*, 4(2974):1–10, 2013. doi:10.1038/ncomms3974.
- [243] J. Fang and H. Li. A facile way to tune mechanical properties of artificial elastomeric proteins-based hydrogels. *Langmuir*, 28(21):8260–8265, 2012. doi:10.1021/la301225w.
- [244] H. Li and Y. Cao. Protein mechanics: from single molecules to functional biomaterials. *Acc. Chem. Res.*, 43(10):1331–1341, 2010. doi:10.1021/ar100057a.
- [245] D. J. Sirbulu, R.W. Friddle, J. Villanueva, and Q. Huang. Nanomechanical force transducers for biomolecular and intracellular measurements: is there room to shrink and why do it? *Reports Prog. Phys.*, 78(2):024101(1)–024101(22), 2015. doi:10.1088/0034-4885/78/2/024101.
- [246] Michael T Woodside and Steven M Block. Reconstructing folding energy landscapes by single-molecule force spectroscopy. *Annu. Rev. Biophys.*, 43:19–39, 2014. doi:10.1146/annurev-biophys-051013-022754.
- [247] F. Benedetti, C. Micheletti, G. Bussi, S. K. Sekatskii, and G. Dietler. Nonkinetic modeling of the mechanical unfolding of multimodular proteins: theory and experiments. *Biophys. J.*, 101(6):1504–1512, 2011. doi:10.1016/j.bpj.2011.07.047.
- [248] T. Bornschlöggl and M. Rief. Single-molecule dynamics of mechanical coiled-coil unzipping. *Langmuir*, 24(4):1338–1342, 2008. doi:10.1021/la7023567.
- [249] Z. N. Scholl, Q. Li, and P. E. Marszalek. Single molecule mechanical manipulation for studying biological properties of proteins, DNA, and sugars. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.*, 6(3):211–229, 2014. doi:10.1002/wnan.1253.

- [250] F. Berkemeier, M. Bertz, S. Xiao, N. Pinotsis, M. Wilmanns, F. Gräter, and M. Rief. Fast-folding alpha-helices as reversible strain absorbers in the muscle protein myomesin. *Proc. Natl. Acad. Sci. U. S. A.*, 108(34):14139–14144, 2011. doi:10.1073/pnas.1105734108.
- [251] A. E. M. Beedle, A. Lezamiz, G. Stirnemann, and S. Garcia-Manyes. The mechanochemistry of copper reports on the directionality of unfolding in model cupredoxin proteins. *Nat. Commun.*, 6(7894):1–9, 2015. doi:10.1038/ncomms8894.
- [252] J. Perales-Calvo, A. Lezamiz, and S. Garcia-Manyes. The mechanochemistry of a structural zinc finger. *J. Phys. Chem. Lett.*, 6(17):3335–3340, 2015. doi:10.1021/acs.jpcllett.5b01371.