

promoting access to White Rose research papers



Universities of Leeds, Sheffield and York
<http://eprints.whiterose.ac.uk/>

This is the author's post-print version of an article published in **Physical Chemistry Chemical Physics, 15 (38)**

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/id/eprint/77850>

Published article:

Hoffmann, T, Tych, KM, Hughes, ML, Brockwell, DJ and Dougan, L (2013)
Towards design principles for determining the mechanical stability of proteins.
Physical Chemistry Chemical Physics, 15 (38). 15767 - 15780. ISSN 1463-9076

<http://dx.doi.org/10.1039/c3cp52142g>

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Towards design principles for determining the mechanical stability of proteins

Toni Hoffmann^{a,b†}, Katarzyna M. Tych^{a,b†}, David J. Brockwell^b and Lorna Dougan^{a,b,*}^a School of Physics and Astronomy, University of Leeds, Leeds, United Kingdom LS2 9JT^b Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, United Kingdom

Received (in XXX, XXX) Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXXX 20XX

DOI: 10.1039/b000000x

10

The successful integration of proteins into bionanomaterials with specific and desired function requires an accurate understanding of their material properties. Two such important properties are their mechanical stability and malleability. While single molecule manipulation techniques now routinely provide access to these, there is a need to move towards predictive tools that can rationally identify proteins with desired material properties. We provide a comprehensive review of the available experimental data on the single molecule characterisation of proteins using the atomic force microscope. We uncover a number of empirical relationships between the measured mechanical stability of a protein and its malleability which provide a set of simple tools which might be employed to estimate properties of previously uncharacterised proteins.

1. Introduction

Proteins are biological nanomachines that utilise mechanical forces in a wide range of cellular processes¹⁻⁴. These important processes range from the translocation of proteins/DNA across membranes^{5, 6}, the degradation of proteins by molecular chaperone proteins⁷, the mechanical resilience of proteins within a molecular scaffold⁸⁻¹¹ and the conversion of mechanical signals into electrochemical signals^{12, 13} (Figure 1). In isolation or as a component of larger complexes, proteins perform their function through structural changes, modifying their intra- and intermolecular interactions. The folded, native conformation of a protein is stabilised by weak localised interactions including electrostatic interactions, van der Waals forces, hydrogen bonds and the hydrophobic effect¹⁴. These same forces are also important in stabilising intermolecular bonds in protein complexes.

The native conformation of a protein represents a minimum of its free energy. Protein stability is only marginal as their free energies of unfolding range from 5 to 15 kcal mol⁻¹ (8-25 k_BT)^{15,16}. Changes in protein conformation upon unfolding are measured in the nanometre length scale, and given the energies

involved, the relevant biological forces are expected to be in the piconewton range. Proteins are subject to thermal forces and the number of possible configurations of the protein (entropy) is at its maximum when it forms a random coil or is denatured. This entropy is reduced as the protein forms secondary or tertiary structures. Extending these native tertiary structures, to overcome the forces holding them together, has been achieved experimentally using a number of single molecule manipulation techniques and requires forces of the order of piconewtons¹⁷. More than a decade ago, a pioneering study used an instrument called an atomic force microscope (AFM) to mechanically unravel a single molecule of the muscle protein titin¹⁰. This study showed that the protein exhibited resistance to unfolding, with forces of 150 – 350 pN being required to unravel the molecule. The mechanical stability, F_U , or resistance to unfolding in response to an applied mechanical force, is therefore a parameter of physiological importance, allowing a molecule to remain folded under certain mechanical stress. The malleability of a protein is a measure of its ability to be deformed without breaking or unfolding.

As well as the study of biological proteins and their role *in vivo*, significant advances have been made in the study and use of proteins in the rational design of new, materials¹⁸⁻³². For example, spider silk proteins have been examined in detail to determine the relationship between protein sequence, structure and material properties. This approach promises a path towards the next generation of bio-materials for mechanically robust applications³³. Elastomeric proteins act as important functional units in biomechanical machinery. These proteins are now beginning to be exploited as the building blocks for biological materials that exhibit outstanding mechanical properties, as they possess the desired elasticity, mechanical strength and resilience required for these functional materials. Inspired by the muscle protein titin, synthetic multidomain polymers have been developed, in the pursuit of materials with combined mechanical properties of mechanical strength and elasticity³⁴⁻³⁷. These studies demonstrate the importance of non-covalent, intramolecular interactions in achieving advanced mechanical properties for

proteins and biomimetic polymers. Recent studies have shown examples of engineered elastomeric proteins with mechanical properties that mimic and surpass those of natural elastomeric

proteins²⁷, and have utilised natural elastomeric proteins that are well-characterised on the nano-scale to engineer hydrogels with specific macro-scale mechanical properties²⁶ (Figure 2A).

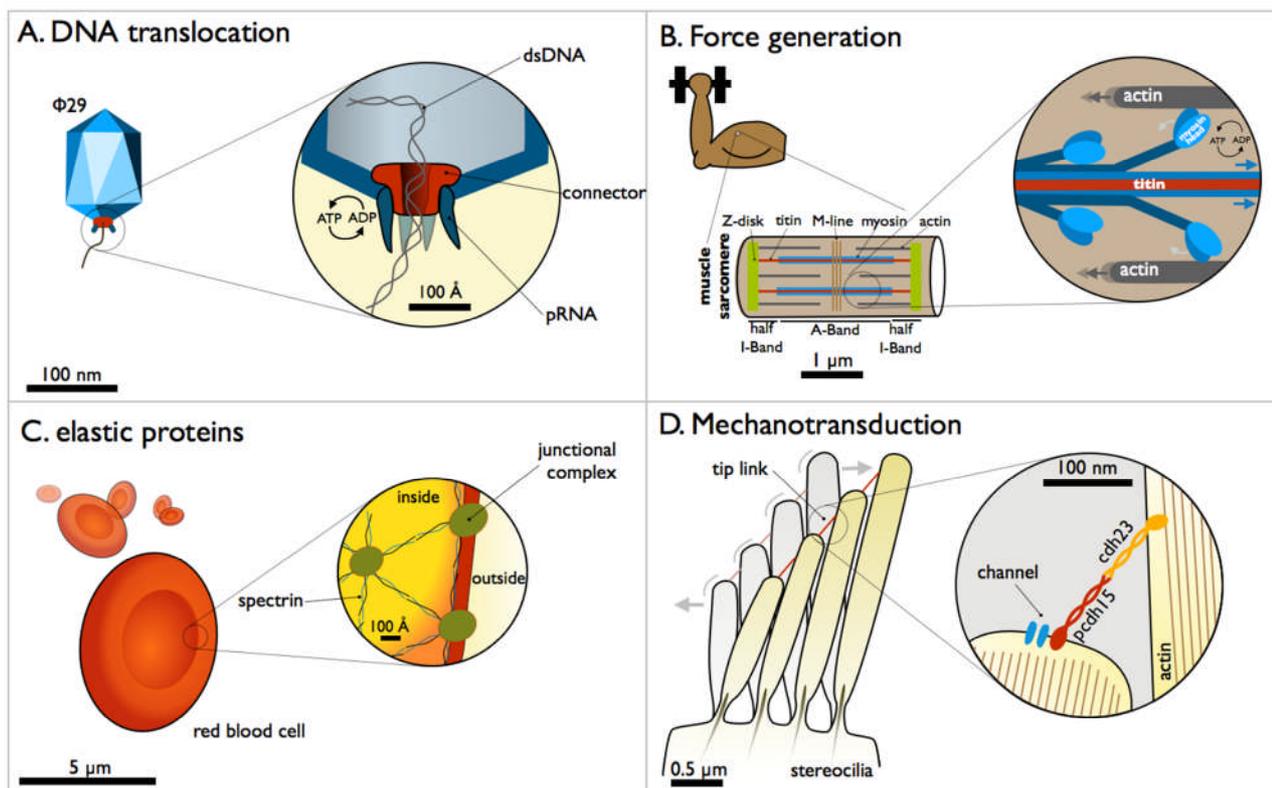


Fig. 1 Proteins as biological machines *in vivo*. (A) DNA translocation is required during packaging of a viral genome into the preformed protein hull (procapsid). In the bacteriophage $\Phi 29$ it is accomplished by one of the strongest known biological motors. (B) Force generation in muscles happens within the repeating units of muscle sarcomeres. It relies on the cyclical binding of the myosin filaments to adjacent actin filaments. A conformational change powered by the hydrolysis of ATP pulls the actin filaments inwards and consequently shortens the sarcomere. With this motion, the muscle contracts. (C) Proteins are functionally important due to their resistance to mechanical force. For example, the α -helical protein spectrin and other proteins form the elastic network of the cytoskeleton. It gives red blood cells their unique flow-optimised shape and their elastic properties. (D) Proteins convert mechanical signals into electrochemical signals. The sensory cells in the inner ear of mammals are equipped with bundles of large membrane-covered cell protrusions, so called stereocilia. Stereocilia pivot when they are mechanically stimulated by sound. The tips of the cilia are linked by protein tethers made of cadherin 22 and protocadherin 15. This tip link is anchored within the membrane to an ion channel. A deflection of the stereocilia opens and closes the ion channels that results in changes of ion flux across the membrane. A sufficiently strong deflection will eventually depolarise the cell and lead to an electrical potential that can reach the auditory nerve

Another study exploited the architecture found in spider silk proteins to engineer materials with remarkable extensibility and strength²⁹ (Figure 2B). The use of proteins in the rational design of biomimetic materials and functional biomaterials for tissue engineering, lubrication and medicine, is now a field of considerable and growing current interest^{23, 28, 38}. To exploit proteins for the design of artificial, novel materials or to utilise them in nanomechanical systems as springs, switches or sensors²³, it will be necessary to have a tool-box of proteins available with known or predictable mechanical properties. Although the number of experimentally studied proteins is ever increasing, it is still very limited. Uncovering some of the design principles that underlie protein stability and flexibility is an important step towards achieving that goal. In addressing this challenge, the ability to predict properties such as the mechanical stability, malleability and flexibility of such materials under different environmental conditions is highly desirable. An understanding of their structural characteristics and mechanical

properties from the smallest scale is essential to enable their efficient and full exploitation.

Single molecule force spectroscopy has emerged as a powerful tool to investigate the forces and motions associated with biological molecules. The most common force spectroscopy techniques are optical tweezers, magnetic tweezers and AFM^{17, 39-42}. Here we focus on approach taken using the AFM, which has been used for more than a decade to study the mechanical properties of a broad range of proteins^{10, 43}. This technique is advancing, and the number of natural and designed proteins studied in experiments, combined with those characterised by computational modelling, provide a growing data set for a detailed analysis of the mechanical stability of proteins^{23, 25, 44-49}. It is by using this growing body of data that we seek to establish a set of simple tools, to estimate the parameters that decide the protein unfolding landscape, prior to in-depth experimental characterisation.

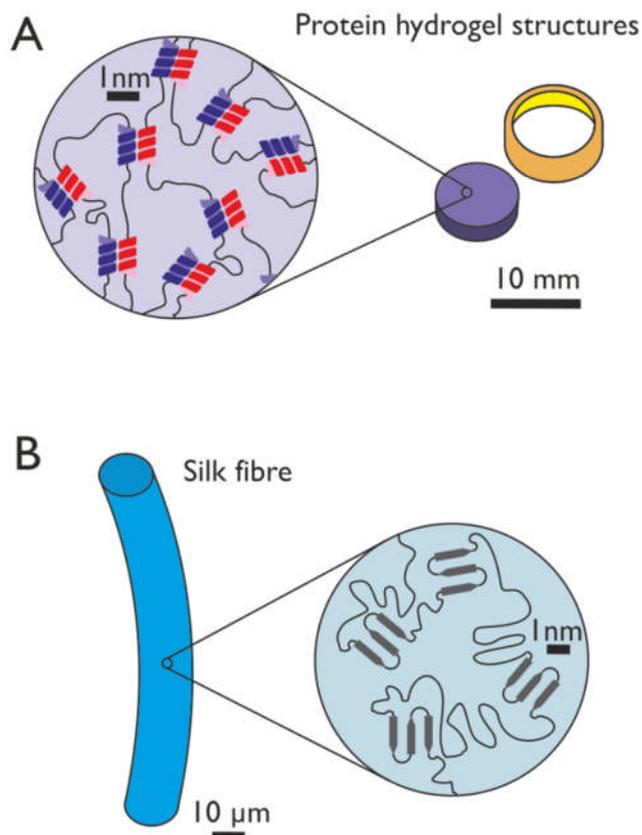


Fig. 2 Proteins as biological modules in the design of new materials (A) A schematic of the protein-based hydrogel described by Lv, Cao and Li²⁶, where the self-assembly of two complementary leucine-zipper sequences into coiled-coils at pH 7.0 is utilised to bond a network of tandem modular proteins. The resulting hydrogels can be engineered to have particular mechanical properties depending on the proteins used to make the networks, and the bond geometries. (B) A schematic of the structure of silk, where beta-sheets act as cross-links joining filaments. The arrangement of these beta-sheets provides silk with its remarkable extensibility and strength²⁹.

In this perspective we aim to provide the non-specialist with a review of the current experimental data characterising the mechanical stability of single proteins as well as providing a viewpoint on the future direction of the field. In section 2 we begin with an introduction to the technique of single molecule force spectroscopy for the study of protein mechanical stability. In section 3 we review the current experimental data available in the literature on protein mechanical stability using the AFM. In section 4 we begin to identify predictive tools for calculating the mechanical stability of proteins and in section 5 and 6 we examine the relationship between protein mechanical stability, malleability, the underlying energy landscape, and protein structure. Finally, we conclude with a summary and thoughts on the future directions of this field.

2. Single molecule force spectroscopy to study proteins

With the advent of single molecule manipulation techniques it is now possible to manipulate single proteins and study their

mechanical properties. The techniques include AFM, optical tweezers, magnetic tweezers and the biomembrane force probes^{17, 39}. Single molecule force spectroscopy, using the atomic force microscope (AFM), is one of the nanomanipulation techniques most extensively used for the study of the mechanical properties of proteins^{44, 50}. In an AFM force extension experiment a protein is extended and unfolded at a constant velocity, yielding information on the mechanical stability of the protein, or the force required to unfold it, F_U . The process is described in detail in Figure 3.

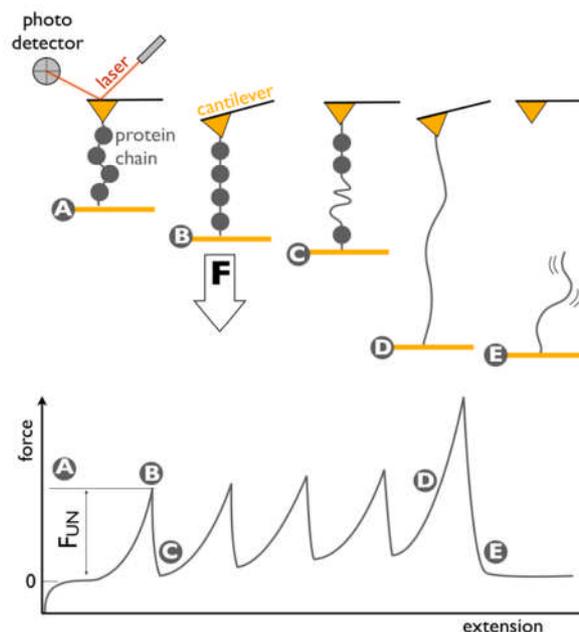


Fig. 3 In single-molecule force spectroscopy, a polyprotein containing repeating protein domains (grey circles) is tethered between the tip of a cantilever and a substrate (A). Increasing the distance between tip and substrate exerts a force on the tethered protein chain (B) which in turn displaces the cantilever (with a known spring constant). By focusing a laser on the back of the cantilever tip, this displacement is detected as a change in the position of the reflected laser light on a photodetector. In force-extension experiments, the protein chain is pulled at constant velocity. The increasing force leads to the subsequent unfolding of single domains within the protein chain which results in a sudden drop in the pulling force (C). The process repeats until all of the domains are unfolded (D). Eventually, the protein will detach from the tip (E). The entire process results in a typical sawtooth-like force extension plot as monitored by the displacement of the AFM cantilever. This plot reveals the force required to unfold (F_U) the domains within the protein chain.

The unfolding of a protein under an external force can be described as a lowering of the free energy barrier between the folded and unfolded state of the protein (Fig. 4). This increases the likelihood of thermal fluctuations leading to a transition from the folded to the unfolded state. For a two-state unfolding process, this reduction in the energy barrier is dependent on the magnitude of the applied force and the distance between the free energy barrier and the folded state energy well, as described by the Bell model⁵¹:

$$k(F) = A \exp \left[\frac{-(\Delta G - F\Delta x_U)}{k_p T} \right] \quad (1)$$

where $k(F)$ is the force-dependent rate constant, F is the applied force, A is the attempt frequency, Δx_U is the distance from the folded state to the transition state, k_B is Boltzmann's constant and T the temperature⁵¹. The value of Δx_U is determined by the distance of the transition state relative to the native, folded state along the unfolding pathway. A movement of the transition state towards the unfolded state will result in an increased Δx_U . Single molecule AFM experiments allow Δx_U and ΔG^* to be quantified, (given an estimate for the exponential pre-factor, A), uncovering features of the underlying energy landscape of proteins⁵²⁻⁵⁵. While the Bell model is most frequently employed to extract information on the unfolding energy landscape of a protein using AFM, it should be noted that a number of alternative models have now been proposed in the literature⁵⁶⁻⁶⁴. We refer the reader to this literature for further information.

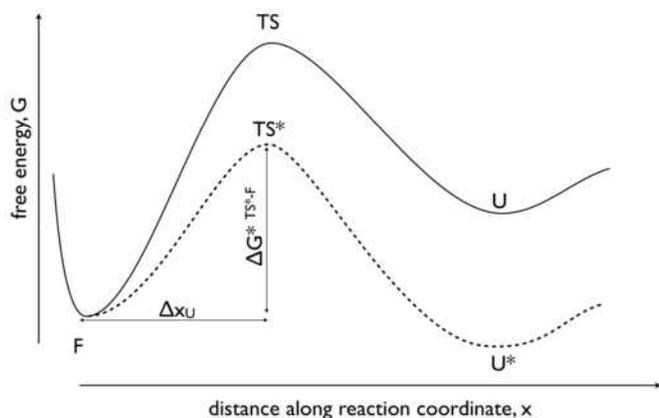


Fig. 4 The unfolding pathway of a protein can be depicted in a free-energy profile. In case of a simple two-state transition, the protein moves from its native, folded state (F) through a transition state (TS) to a denatured, unfolded state (U) at an unfolding rate k_U . The distance between native and transition state is described by Δx_U . The application of an external pulling force causes the free-energy profile to tilt (dotted line). Thus, the energy barrier (ΔG^*) to reach the unfolded state (U^*) is lowered. A sufficiently high force will deform the energy profile such that the unfolded state becomes favoured over the folded state.

In characterising the mechanical stability of a protein, it is common practice to perform single molecule force spectroscopy experiments at several different pulling speeds⁶⁵. Generally, a dataset containing a large number of unfolding forces of a given protein at a single pulling speed is plotted in the form of a histogram, and the median unfolding force value, and a measure of the spread of the data, are obtained (Figures 5 A and B). This is repeated over several different pulling speeds, enabling the dependence of the force on the pulling speed to be plotted (Figure 5 C). It is this pulling speed dependence of the unfolding force, as well as the measure of the width of the unfolding force distribution, that enable the underlying features of the unfolding energy landscape to be extracted from the data using the Bell model (Figure 5D).

A number of proteins have now been studied and their mechanical stability and pulling speed dependence on mechanical stability have been determined. Figure 6 shows examples of the different proteins that have been studied using this approach. More details and references can be found in Table 1. Over the

past decade, a number of studies have contributed towards pinpointing the interactions and structural features of proteins responsible for their mechanical stability⁴⁴. These studies have demonstrated that proteins can be ranked according to their secondary structure content and arrangement – where mostly alpha-helical proteins are mechanically weaker (low F_U) than those predominantly composed of beta-sheet structures (higher F_U)^{44, 46}.

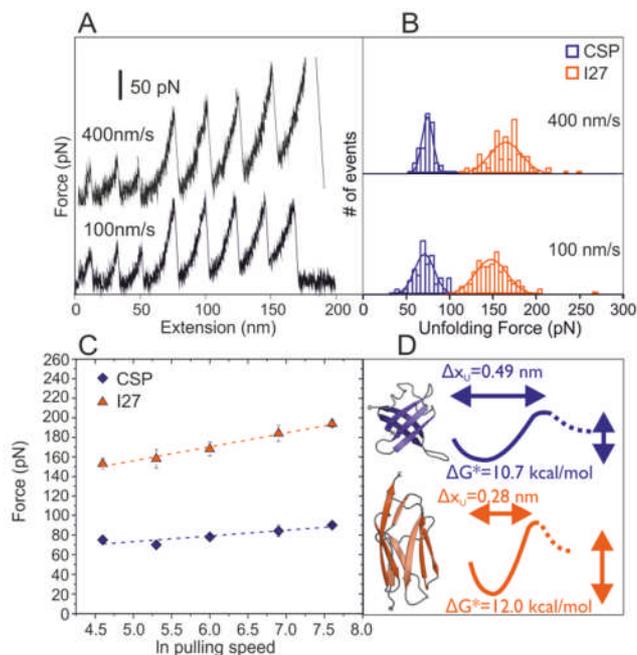


Fig. 5 The effect of pulling speed on the unfolding forces of cold shock protein (CSP) from *Thermotoga maritima*, and I27, adapted from⁴³. A polyprotein construct containing three CSP domains and four I27 domains is unfolded at two different pulling speeds, 100 nm/s and 400 nm/s. (A) The sawtooth patterns resulting from unfolding full polyprotein constructs. (B) The unfolding force distributions with Gaussian fits to obtain a measure of the spread of the data. (C) The median unfolding forces for CSP (blue circles) and I27 (orange squares) plotted against the natural logarithms of the pulling speeds (data not shown). (D) Energy landscapes estimated using the Bell model, and three-dimensional ribbon representation structures of I27 (orange) and CSP (blue). The PDB accession codes for the structures are I27 - 1TIT, CSP - 1G6P.

The importance of the arrangement of the secondary structure in relation to the direction of the pulling force has been demonstrated, where for example the shearing apart of two beta strands requires a greater force than “un-zipping” them sequentially^{56, 66-69}. Indeed, an early molecular dynamics study on the I27 protein identified a ‘mechanical clamp’ region within the secondary structure which involved two neighbouring beta-strands⁷⁰. Mechanical clamps have since been identified in many other proteins^{9, 71-74}. Further studies have examined side chain packing and long-range interactions in topologically similar proteins⁵², hydrophobic packing in the hydrophobic core of a protein⁷⁵, solvent accessibility of hydrogen bonds⁷⁶, non-native interactions⁷¹ and bond patterns as well as the identification of “strong” and “weak” sequence motifs in protein families^{24, 25, 76-79}.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

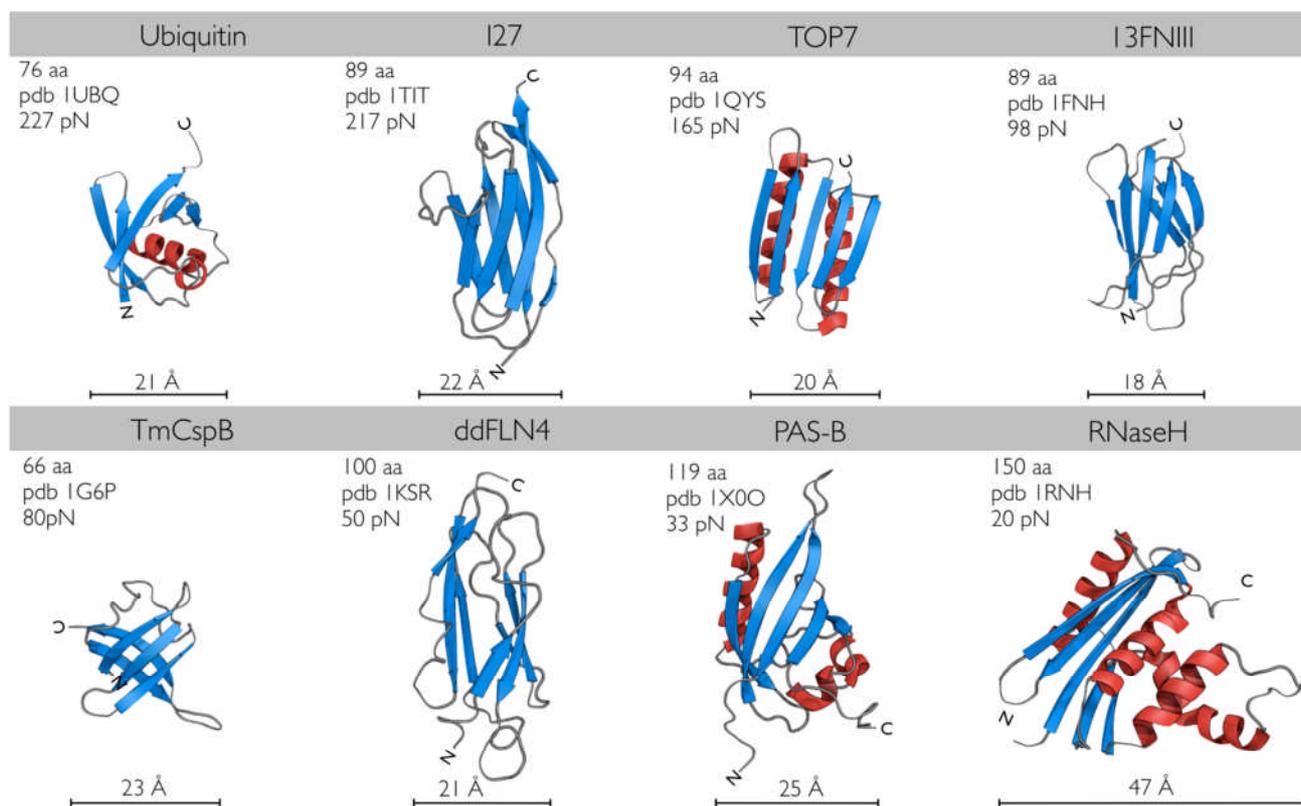


Fig. 6 A selection of proteins studied in single-molecule force-extension experiments. Shown are ribbon representations of the tertiary structure of selected proteins. β -strands are shown as blue arrows and α -helices are represented as red ribbons. The proteins are extended from their amino- and carboxy-terminal ends (N- & C-).

3. Survey of single molecule protein unfolding data

We have completed an extensive survey of the available literature to find all single molecule protein unfolding studies using AFM. From these studies, of which there are many, we found a dataset of 25 proteins for which the pulling speed dependence of the mechanical stability had been determined (see Table 1 & 2 for full details of references). For the current study, we assume that the proteins follow a linear relationship between the unfolding force and pulling speed, in accord with the literature reference from which the data is taken. A linear fit has been applied to each published data set to obtain the unfolding force pulling speed dependence for a range of forces from 100 nm/s to 1000 nm/s. In Figure 7 we show the mechanical stability F_U as a function of pulling speed for a set of 25 different proteins. For the proteins studied to date, it can be seen from Figure 7 that there is a wide range of mechanical stability ranging from low values of tens of

piconewtons for the all alpha helical protein calmodulin (Cam DomC) to high values of hundreds of piconewtons for the all beta sheet protein rubredoxin (Fe-pfRD).

As well as differing values of F_U , the dependence of F_U on pulling speed can also be seen in Figure 7. The gradient of F_U versus the natural logarithm of the pulling speed, gives a measure of the mechanical sensitivity of the protein to the speed at which it is unfolded by force. Mechanically strong proteins such as rubredoxin (Fe-pfRD) exhibit a steep gradient, while mechanically weak proteins such as calmodulin (Cam DomC) exhibit a shallow gradient.

To quantitatively compare the mechanical sensitivity of all 25 proteins in Figure 7, we calculated the gradient of the speed dependence of the unfolding force for each protein at a given pulling speed. In Figure 8 we show the gradient ($\Delta F_U / \Delta \ln(v)$) versus the measured mechanical stability F_U of each protein at a pulling speed of 600 nm/s. For all 25 proteins we find a positive correlation between the magnitude of the mechanical stability F_U and the change in mechanical stability with pulling speed ($\Delta F_U / \Delta \ln(v)$), with mechanically strong proteins (high F_U) having

a large value for the $\Delta F_U/\Delta \ln(v)$, and mechanically weak proteins (low F_U) having a small value for the gradient ($\Delta F_U/\Delta \ln(v)$). Given the gradient ($\Delta F_U/\Delta \ln(v)$) is a measure of the force

sensitivity of the protein to unfolding speed, this suggests that mechanically strong proteins are more force-sensitive than mechanically weak proteins.

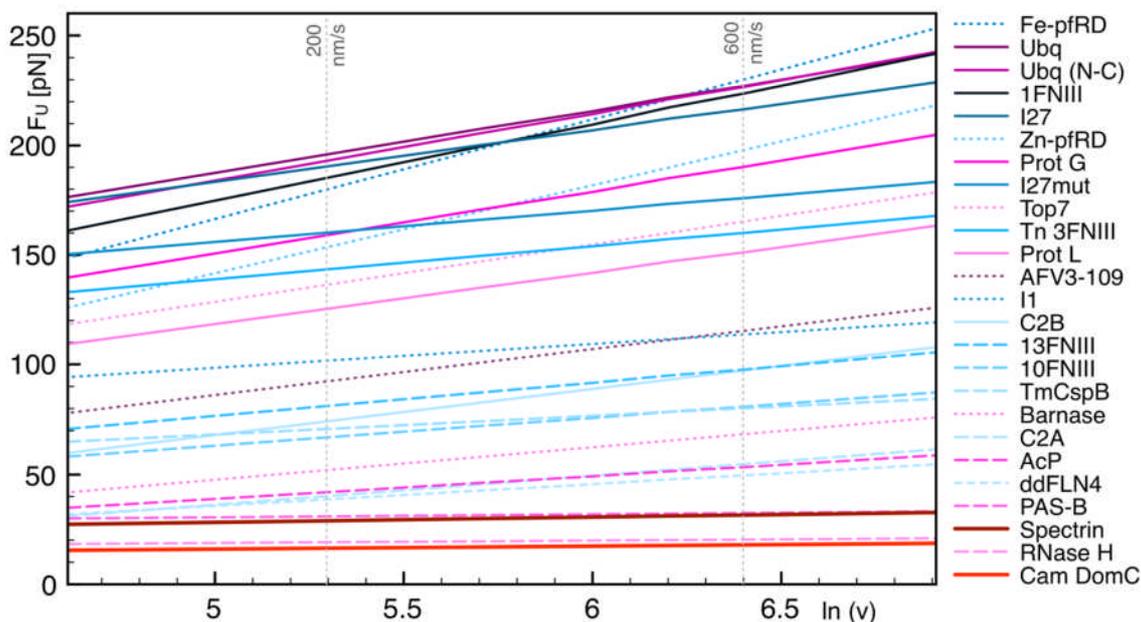


Fig. 7 The experimental pulling speed dependence of the unfolding force of a protein (F_U) for 25 proteins from the literature (Table 1). A linear fit has been applied to each published data set to obtain this plot. Published data that is only provided as a graph was extracted and converted into numerical values using the (x,y) pixel coordinates of individual data points. The protein labels are given on the right in decreasing order regarding their respective unfolding forces at a pulling speed of 600 nm/s. The color scheme indicates all beta proteins in shades of blue, proteins with mixed beta-sheet/alpha-helical content in purple hues, and pure alpha-helical proteins in shades of red.

In summary, by completing AFM force-extension experiments of protein unfolding, the mechanical stability, F_U , of a protein at a range of different pulling speeds can be measured (Figure 7). It is worth noting that these experiments take time, as sufficient statistics need to be gathered to obtain the distributions of unfolding forces (e.g. Figure 5B), and experiments are often completed in triplicate to ensure reliability/robustness³⁹.

While these studies have provided detailed information for specific protein folds, there is a need to move towards more high-throughput, predictive tools for understanding the mechanical stability of proteins as well as the dependence of mechanical stability on pulling speed. As a first step, we have compared the mechanical stability of all 25 proteins in our dataset to determine the relationship between the change in mechanical stability with change in pulling speed ($\Delta F_U/\Delta \ln(v)$) and the measured mechanical stability F_U (Figure 8).

Next, we consider how we can use this information to find a relationship between the mechanical stability F_U and the pulling speed, at all pulling speeds, the experimental variable in AFM experiments. Such a relationship would remove the need to complete a full experimental study of the pulling speed dependence of the force required to unfold a protein, as this information would be accessed by measuring F_U at only one pulling speed.

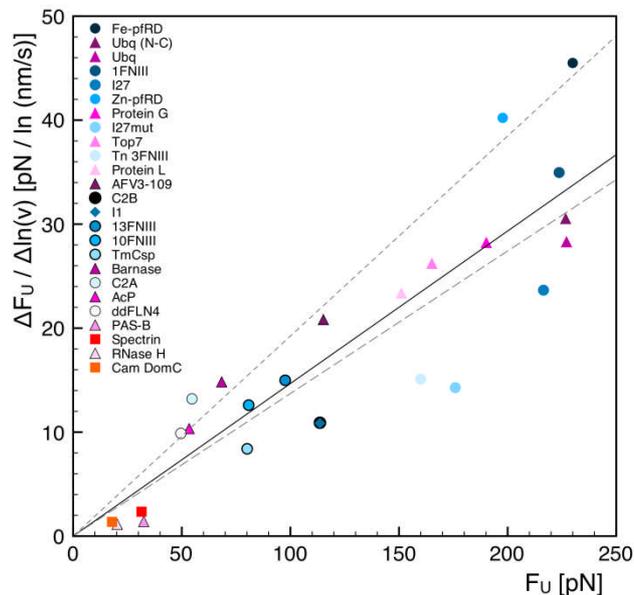


Fig. 8 The dependence of force sensitivity of the unfolding force F_U . The solid line is the linear fit to the available data of 25 experimentally studied proteins pulled at 600 nm/s. Dashed lines indicate how this fit tilts when 100 nm/s (upper dashed line) or 1000 nm/s (lower dashed line), is used. The colour scheme indicates all beta proteins in shades of blue, proteins with mixed beta-sheet/alpha-helical content in purple hues, and pure alpha-helical proteins in shades of red.

4. Towards predictive tools of protein mechanical stability

To access information about the unfolding energy landscape of a protein the unfolding force is measured at a range of different pulling speeds (Figure 5). It would be valuable to have a tool which would allow the dependence of unfolding force on pulling speed to be determined with minimal effort, for example after the completion of one experiment at one unfolding speed. If one unfolding force was experimentally obtained, F_U , at a pulling speed v , we could use the information in Figure 8 to obtain a predicted dependence of unfolding force on pulling speed $\Delta F_U/\Delta \ln(v)$, so that the unfolding force at any pulling speed could be calculated, F' and v' . Here we describe one approach for how this might be achieved.

In Figure 8 we showed that the force sensitivity of a protein to pulling speed ($\Delta F_U/\Delta \ln(v)$) can be related to the protein mechanical stability F_U at a given pulling speed, v . At a pulling speed of 600 nm/s we find that $\Delta F_U/\Delta \ln(v)=0.15F_U$ with an $R^2=0.78$. When plotted for different pulling speeds (Dashed lines in Figure 8), the mechanical stability of the protein changes and as a result a different dependence between $\Delta F_U/\Delta \ln(v)$ and F_U is found. Using the available experimental data (Figure 8) we can find a relationship between the mechanical stability sensitivity of a protein and the pulling speed. We find that $\Delta F_U/\Delta \ln(v)$ and F_U are related by $\Delta F_U/\Delta \ln(v) = 0.7 \ln v^{(-0.84)} F_U$. Therefore, if an experiment is completed at one pulling speed, v , and an unfolding force F_U is obtained, the relationship above can be used to predict $\Delta F_U/\Delta \ln(v)$. By integrating this equation we can find a more general description of F_U . This equation allows us to predict the expected unfolding force (F_U) at a given pulling speed (v) for a protein with a known unfolding force (F'_U) at a single pulling speed (v').

$$F_U = \frac{F'_U}{\exp\left[\frac{0.7}{0.16}[\ln(v')]^{0.16}\right]} \cdot \exp\left[\frac{0.7}{0.16}[\ln(v)]^{0.16}\right] \quad (2)$$

This relationship permits the pulling speed dependence of the protein to be determined for a range of different pulling speeds, and as a result parameters of the energy landscape of the protein could be extracted (Figure 5).

We tested the robustness of this expression by calculating the unfolding forces for all 25 proteins shown in Figure 8, using the unfolding force at 600 nm/s pulling speed as our input in equation 2. We then compared the unfolding forces in Figure 8 with the calculated unfolding forces, within a pulling speed range of 100 to 1000 nm/s, and found the root mean squared error (RMSE). The RMSE are shown in Table 1 for each of the proteins, where a low RMSE value indicates that the deviation from experimental and calculated forces is low.

To exploit the expression further we have implemented this relationship for a number of different proteins in the literature that have been mechanically unfolded using the AFM, at one pulling speed (Table 2). Using equation (2) we have predicted the

pulling speed dependence of the unfolding force for six proteins from the literature, which have been studied at just one pulling speed (Figure 9). The all-alpha-helical protein vascular cell adhesion molecule-1 (VCAM1) has previously been studied using AFM force extension at a pulling speed of 1000 nm/s, measuring an unfolding force of 40 pN⁸⁰. We predict that the mechanical stability will change from 27 pN – 40 pN as the pulling speed is increased from 100 nm/s to 1000 nm/s. The pure beta sheet single cohesin domain from the scaffolding protein CipA (scaffoldin c7A) has been mechanically unfolded at 400 nm/s, yielding a mechanical stability of 480 pN⁷⁴. We predict that this mechanically strong protein will unfold at a force of between 379 – 549 pN as the pulling speed is increased from 100 to 1000 nm/s. Thus this protein is mechanically very strong (the strongest measured to date) and if the pulling speed dependence we predict were experimentally confirmed, the scaffoldin c7A protein would have the highest sensitivity in mechanical stability towards an applied pulling speed ($\Delta F_U/\Delta \ln(v)$).

Clearly, this model cannot predict atypical behaviour between proteins with very similar unfolding forces as it only incorporates the average behaviour of many different proteins, for example it cannot take into account the effect of a proline mutation in the mechanical clamp region of the protein⁸¹. Nor does the method account for any deviations from non-linear behaviour in the unfolding force. As such it gives a global perspective on the range of experimentally explored force sensitivities. We propose that it may be useful for rapidly assessing how a protein deviates from the average observed dependence of the unfolding force from the pulling speed. Aberrant behaviour could point towards unusual topologies or molecular interactions that modify the proteins response to an applied force. Moreover, the model could serve as a useful template that allows the integration of modulating factors that affect the mechanical stability and force sensitivity of a protein.

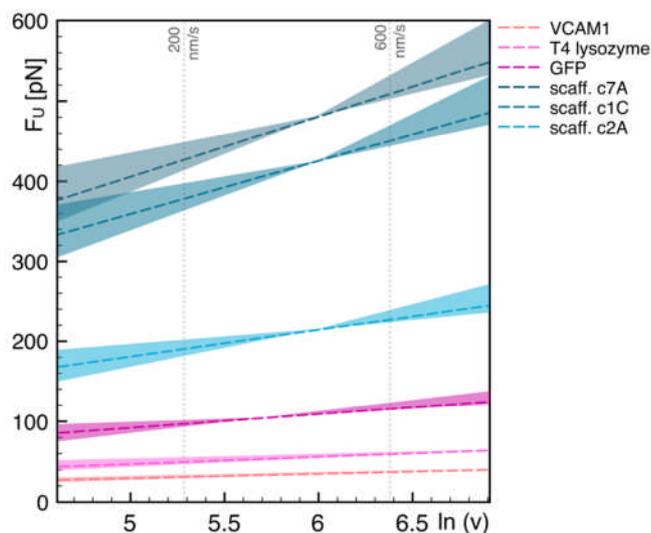


Fig. 9 Best fit estimations for the dependence of the unfolding force from the pulling speed (between 100 - 1000 nm/s) for proteins where only a single pulling speed has been published. Shaded areas indicate the range of possible values based on the root mean square value between experimental and predicted fits for proteins with known dependence of the unfolding force from the pulling speed. The given

colour scheme indicates all beta proteins in shades of blue, proteins with mixed beta-sheet/alpha-helical content in purple hues, and pure alpha-helical proteins in shades of red.

5. Relationship between protein mechanical stability and malleability

Single molecule manipulations techniques have helped to gain insight into the structural bases of protein resistance to forced unfolding, yielding information on mechanical stability F_U and malleability, as measured by Δx_U (Figure 4). A survey of the current experimental literature on the mechanical unfolding of proteins allows us to examine the relationship between F_U and Δx_U (Figure 10). The data shows a robust correlation between F_U and Δx_U with mechanically strong proteins (large F_U) having a small Δx_U , and mechanically weak proteins (small F_U) a large Δx_U . The tendency for all alpha proteins to be mechanically weaker than proteins with mixed alpha-helical / beta-sheet content and pure beta proteins⁸² can be seen in Figure 10. One important development in the understanding of which structural

elements provide mechanical resistance has been the identification of a so-called ‘mechanical clamp’ in many proteins. A mechanical clamp is 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.

This important structural feature is often, but not exclusively, formed between neighbouring β -strands connected by hydrogen bonds. One prominent example can be found in the I27 immunoglobulin-like domain of titin, where the two terminal beta-strands must be sheared apart before the rest of the domain can unfold^{10, 70}. Proteins with mechanical clamp motifs with more complex topology have also been designed *de novo*⁷⁹.

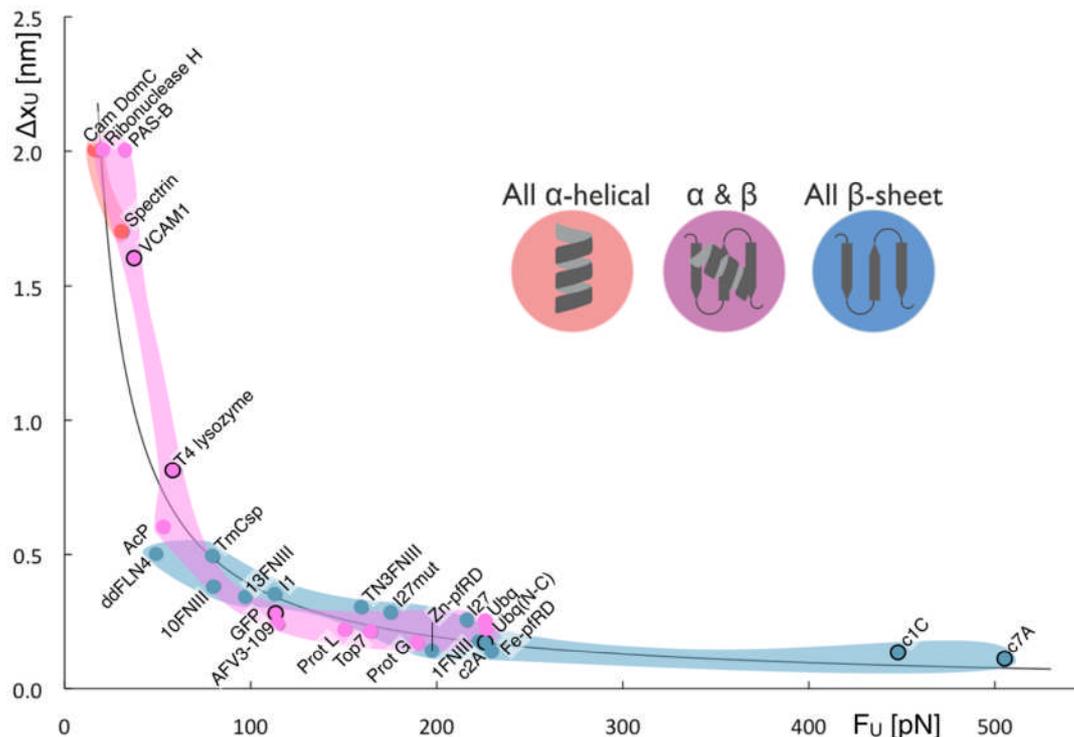


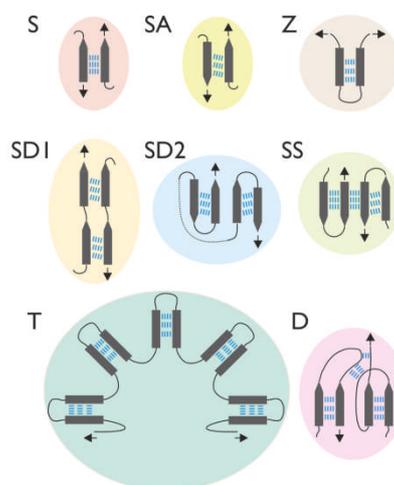
Fig. 10 The relationship between Δx_U and F_U . The unfolding force F_U and the unfolding distance Δx_U are shown for 22 experimentally studied proteins that were unfolded at at least two different speeds. Where required, the expected unfolding force at 600 nm/s was interpolated. The data can be described by a bootstrapped non-linear fit following a power law with $\Delta x_U = 39.4 \pm 3.1 / F_U$ ($R^2 = 0.91 \pm 0.01$). Proteins are grouped according to their general protein category (SCOP): all alpha-helical (red), mixed beta-sheet/ alpha-helical (purple) and pure beta-sheet proteins (blue). Pointed grey arrows depict beta-strands while zigzagged light and dark-grey rectangles illustrate alpha-helices. Five proteins are shown encircled in black. They have a published Δx_U but no given dependence of the unfolding force on the pulling velocity. Here, the expected unfolding force at 600 nm/s has been estimated using the relationship given in Figure 5. Accordingly they have not been used for the power law fit shown in this figure.

A systematic theoretical study of protein secondary structures from the protein data bank permitted the identification of a number of mechanical clamp motifs⁷², defined in Figure 11. 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 in different modifications across all branches of life^{72, 83}. In addition to the mechanical clamp, other studies have examined mechanical

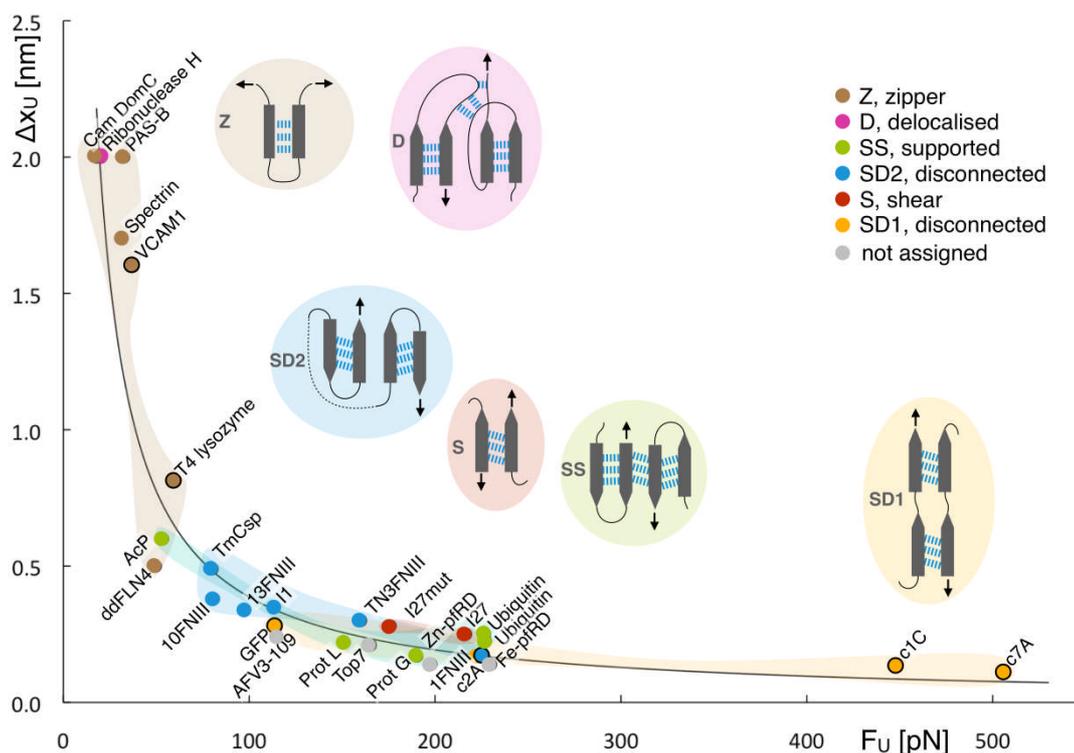
stability by investigating mechanical networks of hydrogen bonds in proteins, mechanical crack propagation and mechanical fracture in the context of protein unfolding under force^{18, 19, 84, 85}.

5 In Figure 12 we show the 25 proteins from our dataset, where we have assigned each protein with a mechanical clamp motif based on the classification system described by Sikora *et al.*⁷². It can be seen that by grouping each of the mechanical clamp motifs, an interesting trend of mechanical hierarchy is observed.
 10 Proteins with a 'zipper' motif, where hydrogen bonds are broken sequentially, exhibit a large Δx_U and small F_U . This suggests that this motif represents proteins that are malleable but mechanically less stable.



15 **Fig. 11 Mechanical clamp motifs according to Sikora *et al.*⁷³. The following abbreviations are used: S ... shear, SA ... shear antiparallel, Z ... zipper, SD1 ... shear disconnected 1, SD2 ... shear disconnected 2, SS ... shear supported, T ... torsion, D ... shear delocalised.**

20



25 **Fig. 12 The relationship between Δx_U and F_U at a pulling speed of 600 nm/s. Schematics as in Figure 7 but proteins are now grouped according to their type of mechanical clamp as described in Figure 10. The data can be described by a non-linear fit following a power law with $\Delta x_U = 39.4 \pm 3.1 / F_U$ ($R^2 = 0.91 \pm 0.01$).**

Proteins with SD1 motifs, where hydrogen bonds must be sheared apart, exhibit a small Δx_U and a broad range of F_U values, implying that this motif provides some malleability as well as versatility in mechanical stability. This figure demonstrates that
 30 most clamp motifs are not yet represented by a large set of experimentally studied proteins. However, among those

characterised experimentally, proteins with a mechanical clamp of the shear-disconnected II type (SD2, Fig. 11) form the largest group. Studies on this motif to date have included the
 35 hyperthermophilic cold shock protein from *Thermotoga maritima*³⁹ and several homologous proteins (fibronectin type III domains and the I1 domain from human cardiac titin), and exhibit

unfolding forces in the range from 70 to 230 pN. Whilst there is some clustering of mechanical clamp motifs in Figure 12, and of secondary structure content in Figure 10 demonstrating that common structural features have an impact on the resulting values of Δx_U and F_U , neither the secondary structure content nor the mechanical clamp motif alone can be used to accurately predict the unfolding force of a protein.

Previous work has suggested that Δx_U is related to the force required to unfold a protein, F_U , by either a power law or a linear correlation⁸². Studies undertaken in the past six years have about doubled the number of proteins with an experimentally determined Δx_U and F_U , allowing us to refine the dependency of Δx_U and F_U and confirm that the relationship is best described by a power law (solid line, Figures 10 and 12) of the form

$$\Delta x_U = \frac{39.4}{F_U} \quad (3)$$

with a chi-squared value of 0.91. In contrast, a linear fit gives a chi-squared value of 0.41. This scaling law indicates that mechanically weaker proteins would have a larger value of Δx_U . Past studies have proposed that an increase in Δx_U represented softening of the protein i.e. a decrease of its spring constant, whereby a protein could be deformed by a greater amount before reaching the transition state and unfolding^{86, 87}. Conversely a protein with a low Δx_U can only be deformed by a small amount before unfolding. Therefore the magnitude of Δx_U can be used as a measure of the deformability or malleability of the protein.

6. Protein mechanical stability and energy landscape

Another parameter that can give insight into the energy landscape of a protein is the product of its F_U and Δx_U . The product reflects the work that is done over the unfolding distance before a protein fold is disrupted under an applied external force (Figure 4). It relates to the energy required to unfold a protein under an applied external force, ΔG^* , and is a measure of the change in the height of the energy barrier between native and unfolding state under an applied force. Clearly, this product of the unfolding force and the distance from the native to the transition state is related to the unfolding rate, k_U at zero force. Proteins with a lower $F_U \cdot \Delta x_U$ unfold faster than proteins with a higher difference in unfolding energy (Figure 13). A lower energy barrier that increases the probability of unfolding might explain this observation, as this would result in faster unfolding. Interestingly, there is no clear correlation between $\log k_U$ and F_U , nor is there one between $\log k_U$ and Δx_U (data not shown). The clear correlation between $\log k_U$ and the $F_U \cdot \Delta x_U$ product highlights the relationship between Δx_U and F_U as two of the major parameters that describe the underlying energy landscape and how they are linked to e.g. the observed unfolding rate of a protein under given experimental conditions.

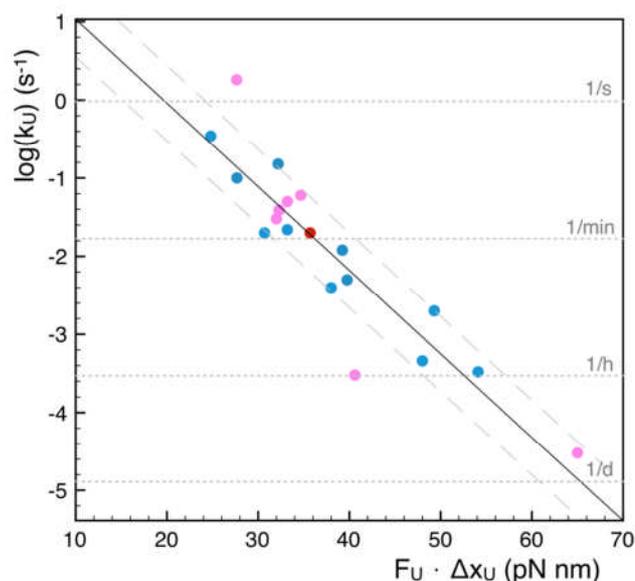


Fig. 13 Dependence of the unfolding rate k_U on the product of $F_U \cdot \Delta x_U$. Proteins are coloured according to their major protein category as either all beta-sheet proteins (blue), all alpha-helical proteins (red) or mixed alpha helical/beta-sheet proteins (purple). Proteins with a high $F_U \cdot \Delta x_U$ unfold several magnitudes of order more slowly than proteins with a low $F_U \cdot \Delta x_U$. The bootstrapped linear fit to the data is given by $\log(k_U) = (-0.107F_U \Delta x_U + 2.1)$ with an ($R^2 = 0.83 \pm 0.02$). Dashed lines indicate the area of the root mean squared error across all bootstrapped fits.

7. Relationship between protein structure and mechanical stability

An interesting parameter to quantify the topology of a protein is its relative contact order (RCO). It is defined as the average sequence distance between all contacting residues normalised to the total length of the protein chain (Fig. 14A)⁸⁸. A low correlation between RCO and the force F_U required to unfold a protein has been reported previously⁸². In our larger data set no clear correlation can be observed between F_U and RCO (Fig. 14A), rather a general trend that a higher RCO leads to a higher unfolding force, in agreement with that reported previously⁸². However, when the studied protein structures are grouped by secondary structure content, it can be seen that all-beta proteins tend to be mechanically more stable when they possess a high RCO (Fig. 14B). No clear relationship is observed neither for mixed alpha-helical/beta-sheet proteins nor for proteins with SD2 or zipper mechanical clamp motifs (as one of the two more frequent motifs in the database). This further demonstrates that while some of the structural features which govern the mechanical stability of protein domains are understood, a selection of different criteria and tools need to be applied in order to be able to predict the behaviour of proteins in response to applied mechanical forces more quantitatively. The poor correlation may partly be due to the insensitivity of contact order to the known anisotropic behaviour of proteins under force. This anisotropy arises as a consequence of the action of mechanical perturbation as local rather global denaturant.

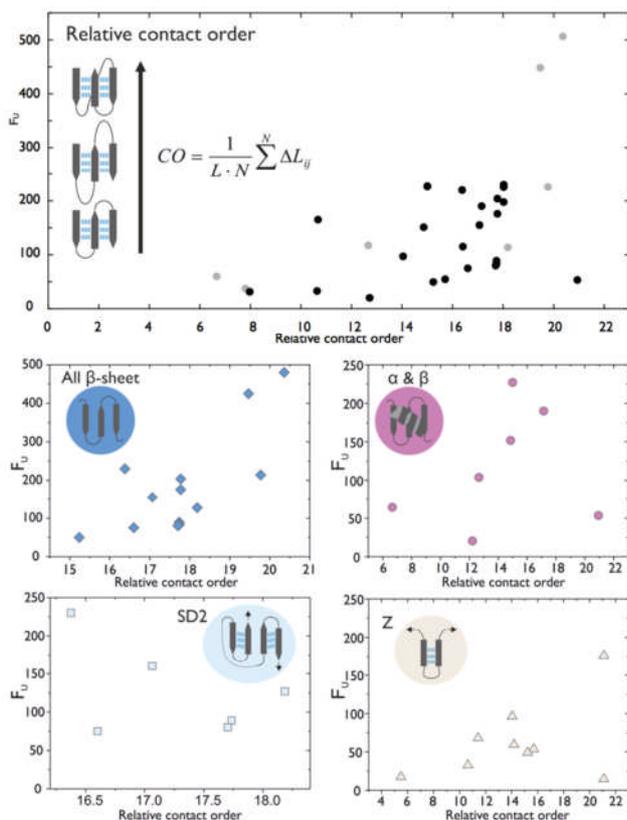


Fig 14. The relationship between the unfolding force (F_U) and relative contact order (RCO). RCO is defined as the average sequence distance between all contacting residues normalised to the total length of the protein chain. In general, proteins with a high RCO tend to be mechanically stronger (Panel A) which is also true for mainly beta proteins. Neither a clear correlation can be observed when proteins of mixed alpha-helical/beta-sheet content are selected, nor for proteins with a SD2 or Zipper mechanical clamp motif.

8. Conclusions

Using single-molecule force spectroscopy we can gain access to the properties of a protein that are relevant from an engineer's perspective such as its mechanical stability and malleability. Such parameters are increasingly important for the rational design of novel, protein-based materials for future applications. Hence, there is a need to move towards predictive tools that can rationally identify target proteins with specific mechanical properties. Only now has the available experimental data set grown to a size that we can start to address questions to uncover common design principles across different types of proteins.

Here, we provide a basic toolbox of correlations that permits the estimation of three important parameters of a protein, the unfolding force (F_U) at an unmeasured pulling speed, the distance to the transition state (Δx_U) and the unfolding rate at force zero

(k_U). We show a relation between force sensitivity and mechanical stability i.e. the dependence of the unfolding force from the applied pulling speed. This enables an estimation of this dependence before further time-intensive experiments are done. Moreover, we report the consolidation of the power law correlation of F_U and the distance to the unfolding state Δx_U . With it, we provide an updated equation that allows a good estimation of Δx_U , a measure for the flexibility of a protein. This relation 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. Moreover, plotting F_U against Δx_U reveals gaps in the explored space of mechanical properties of studied proteins, which will be helpful for the selection of proteins for future force spectroscopy studies. For example, there is a lack of studies on mechanically very strong proteins with unfolding forces above 230 pN and weak proteins below 50 pN at 600nm/s. Finally, an equation for the estimation of the unfolding rate k_U in dependence of Δx_U and F_U is given.

The described correlations represent the average behaviour of many different proteins. They cannot predict the deviating behaviour between variants of proteins previously described or the effect of mutations on the mechanical stability of proteins. However, the correlations can serve as a useful tool to judge how much a studied protein deviates from the average observed behaviour to point out unusual topologies or intramolecular interactions that can modulate the mechanical properties of a protein.

This survey raises further questions such as do all proteins follow this power law relation? Do proteins exist that combine high mechanical stability with high malleability? What are the extreme limits of the mechanical properties of a peptide chain? In this context, the increasing number of studied proteins provides a repository for the selection of protein domains as building blocks to design protein-based materials with desired properties. It has been shown using muscle-mimetic protein polymers^{27, 34-37, 89} that one can combine properties of different proteins that translate to the macroscopic level of protein-based materials. Any rational design of such a material could also take advantage of the observed force anisotropy⁶⁶ to create materials that behave differently depending on the direction of an applied mechanical stress. To address the questions above and to extend the diversity of a repository of building blocks for protein-based materials requires further studies of many proteins preferentially with extreme properties and topologies different to already examined proteins. We hope that this recent survey of available data on mechanically studied proteins together with available databases of simulated protein stretching^{73, 90} will provide a useful overview to guide future studies in this exciting field of research.

† These authors contributed equally

Notes and references

^a School of Physics and Astronomy, University of Leeds, United Kingdom, LS2 9JT. Tel: +44 113 343 8958; E-mail: L.Dougan@leeds.ac.uk

Dr. Lorna Dougan is supported by a grant from the European Research Council (258259-EXTREME BIOPHYSICS). We thank all members of the Dougan group for useful discussions

and feedback, in particular Megan Hughes.

1. M. Strong, in *PLoS Biology*, 2004, vol. 2.
2. F. Berkemeier, M. Bertz, S. B. Xiao, N. Pinotsis, M. Wilmanns, F. Gräter and M. Rief, *P Natl Acad Sci USA*, 2011, **108**, 14139-14144.
3. E. Miller, T. Garcia, S. Hultgren and A. F. Oberhauser, *Biophys J*, 2006, **91**, 3848-3856.
4. B. Bullard, T. Garcia, V. Benes, M. C. Leake, W. A. Linke and A. F. Oberhauser, *P Natl Acad Sci USA*, 2006, **103**, 4451-4456.
5. S. Huang, K. S. Ratliff and A. Matouschek, *Nature structural biology*, 2002, **9**, 301-307.
6. S. Huang, K. S. Ratliff, M. P. Schwartz, J. M. Spenner and A. Matouschek, *Nature structural biology*, 1999, **6**, 1132-1138.
7. R. A. Maillard, G. Chistol, M. Sen, M. Righini, J. Tan, C. M. Kaiser, C. Hodges, A. Martin and C. Bustamante, *Cell*, 2011, **145**, 459-469.
8. M. Carrion-Vazquez, H. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser and J. M. Fernandez, *Nat. Struct. Mol. Biol.*, 2003, **10**, 738-743.
9. J. R. Forman and J. Clarke, *Curr Opin Struc Biol*, 2007, **17**, 58-66.
10. M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez and H. E. Gaub, *Science*, 1997, **276**, 1109-1112.
11. A. F. Oberhauser, P. E. Marszalek, H. P. Erickson and J. M. Fernandez, *Nature*, 1998, **393**, 181-185.
12. D. P. Corey, *F1000 biology reports*, 2009, **1**, 58.
13. M. Sotomayor, W. A. Weihofen, R. Gaudet and D. P. Corey, *Nature*, 2012, **492**, 128-132.
14. P. Nelson, *Biological Physics*, New York, 2004.
15. C. Scharnagl, M. Reif and J. Friedrich, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 2005, **1749**, 187-213.
16. A. R. Fersht, *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*, New York, 1999.
17. K. C. Neuman and A. Nagy, *Nat Methods*, 2008, **5**, 491-505.
18. M. J. Buehler and Y. C. Yung, *Nat Mater*, 2009, **8**, 175-188.
19. M. J. Buehler and Y. C. Yung, *Hfsp J*, 2010, **4**, 26-40.
20. N. Crampton and D. J. Brockwell, *Curr Opin Struc Biol*, 2010, **20**, 508-517.
21. J. Fang and H. B. Li, *Langmuir*, 2012, **28**, 8260-8265.
22. J. Jancar, J. F. Douglas, F. W. Starr, S. K. Kumar, P. Cassagnau, A. J. Lesser, S. S. Sternstein and M. J. Buehler, *Polymer*, 2010, **51**, 3321-3343.
23. H. Li and Y. Cao, *Acc. Chem. Res.*, 2010, **43**, 1331-1341.
24. H. B. Li, W. A. Linke, A. F. Oberhauser, M. Carrion-Vazquez, J. G. Kerkviliet, H. Lu, P. E. Marszalek and J. M. Fernandez, *Nature*, 2002, **418**, 998-1002.
25. W. Lu, S. S. Negi, A. F. Oberhauser and W. Braun, *Proteins*, 2012, **80**, 1308-1315.
26. S. Lv, T. Bu, J. Kayser, A. Bausch and H. Li, *Acta Biomaterialia*, 2013, **9**, 6481-6491.
27. S. Lv, D. M. Dudek, Y. Cao, M. M. Balamurali, J. Gosline and H. Li, *Nature*, 2010, **465**, 69-73.
28. M. Y. Truong, N. K. Dutta, N. R. Choudhury, M. Kim, C. M. Elvin, A. J. Hill, B. Thierry and K. Vasilev, *Biomaterials*, 2010, **31**, 4434-4446.
29. J. Y. Wong, J. McDonald, M. Taylor-Pinney, D. I. Spivak, D. L. Kaplan and M. J. Buehler, *Nano Today*, 2012, **7**, 488-495.
30. Y. H. Yang, C. Dicko, C. D. Bain, Z. G. Gong, R. M. J. Jacobs, Z. Z. Shao, A. E. Terry and F. Vollrath, *Soft Matter*, 2012, **8**, 9705-9712.
31. T. Zhu and J. Li, *Prog Mater Sci*, 2010, **55**, 710-757.
32. M. Kim, C. C. Wang, F. Benedetti, M. Rabbi, V. Bennett and P. E. Marszalek, *Adv Mater*, 2011, **23**, 5684-+.
33. A. Nova, S. Keten, N. M. Pugno, A. Redaelli and M. J. Buehler, *Nano Letters*, 2010, **10**, 2626-2634.
34. Z. B. Guan, *Polym Int*, 2007, **56**, 467-473.
35. Z. B. Guan, J. T. Roland, J. Z. Bai, S. X. Ma, T. M. McIntire and M. Nguyen, *J Am Chem Soc*, 2004, **126**, 2058-2065.
36. A. M. Kushner and Z. B. Guan, *Angew Chem Int Edit*, 2011, **50**, 9026-9057.
37. A. M. Kushner, J. D. Vossler, G. A. Williams and Z. B. Guan, *J Am Chem Soc*, 2009, **131**, 8766-+.
38. B. Lewandowski, G. De Bo, J. W. Ward, M. Pappmeyer, S. Kuschel, M. J. Aldegunde, P. M. E. Gramlich, D. Heckmann, S. M. Goldup, D. M. D'Souza, A. E. Fernandes and D. A. Leigh, *Science*, 2013, **339**, 189-193.
39. F. Ritort, *J. Phys.: Condens. Matter*, 2006, **18**, R531-R583.
40. S. Weiss, *Science*, 1999, **283**, 1676-1683.
41. H. Clausen-Schaumann, M. Seitz, R. Krautbauer and H. E. Gaub, *Current Opinion in Chemical Biology*, 2000, **4**, 524-530.
42. J. R. Moffitt, Y. R. Chemla, S. B. Smith and C. Bustamante, *Annual Review of Biochemistry*, 2008, **77**, 205-228.
43. T. Hoffmann, K. Tych, D. Brockwell and L. Dougan, *Journal of Physical Chemistry, B.*, 2013, **117**, 1819-1826.
44. T. Hoffmann and L. Dougan, *Chem. Soc. Rev.*, 2012, **41**, 4781-4796.
45. A. Galera-Prat, A. Gómez-Sicilia, A. F. Oberhauser, M. Cieplak and M. Carrión-Vázquez, *Curr. Opin. Struct. Biol.*, 2010, **20**, 63-69.
46. N. Crampton and D. J. Brockwell, *Curr. Opin. Struct. Biol.*, 2010, **20**, 508-517.
47. T. Bornschlog and M. Rief, *Langmuir*, 2008, **24**, 1338-1342.
48. I. Schwaiger, A. Kardinal, M. Schleicher, A. A. Noegel and M. Rief, *Nat. Struct. Mol. Biol.*, 2004, **11**, 81-85.
49. V. Aggarwal, S. R. Kulothungan, S. Rajaram, M. M. Balamurali, R. Varadarajan and S. R. K. Ainavarapu, *Biophys. J.*, 2011, **100**, 481-481.
50. G. Zoldak and M. Rief, *Curr Opin Struc Biol*, 2013, **23**, 48-57.
51. G. I. Bell, *Science*, 1978, **200**, 618-627.
52. D. P. Sadler, E. Petrik, Y. Taniguchi, J. R. Pullen, M. Kawakami, S. E. Radford and D. J. Brockwell, *J Mol Biol*, 2009, **393**, 237-248.
53. S. Garcia-Manyes, L. Dougan, C. L. Badilla, J. Brujic and J. M. Fernandez, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 10534-10539.
54. S. Garcia-Manyes, L. Dougan and J. M. Fernandez, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 10540-10545.
55. L. Dougan, K. R. Ainavarapu, G. Genchev, H. Lu and J. M. Fernandez, *Chemphyschem*, 2008, **9**, 2836-2847.
56. R. B. Best, E. Paci, G. Hummer and O. K. Dudko, *J Phys Chem B*, 2008, **112**, 5968-5976.

57. O. K. Dudko, T. G. W. Graham and R. B. Best, *Phys Rev Lett*, 2011, **107**.
58. O. K. Dudko, G. Hummer and A. Szabo, *Phys Rev Lett*, 2006, **96**.
59. O. K. Dudko, G. Hummer and A. Szabo, *P Natl Acad Sci USA*, 2008, **105**, 15755-15760.
60. M. Schlierf and M. Rief, *Biophys J*, 2006, **90**, L33-L35.
61. Z. T. Yew, M. Schlierf, M. Rief and E. Paci, *Phys Rev E*, 2010, **81**.
62. D. K. West, P. D. Olmsted and E. Paci, *J Chem Phys*, 2006, **125**.
63. D. K. West, E. Paci and P. D. Olmsted, *Phys Rev E*, 2006, **74**.
64. Z. T. Yew, P. D. Olmsted and E. Paci, *Adv Chem Phys*, 2012, **146**, 395-417.
65. R. Rounsevell, J. R. Forman and J. Clarke, *Methods*, 2004, **34**, 100-111.
66. D. J. Brockwell, E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham and S. E. Radford, *Nat. Struct. Mol. Biol.*, 2003, **10**, 731-737.
67. M. Carrion-Vazquez, H. B. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser and J. M. Fernandez, *Nature structural biology*, 2003, **10**, 738-743.
68. H. Dietz and M. Rief, *P Natl Acad Sci USA*, 2006, **103**, 1244-1247.
69. H. Dietz and M. Rief, *Phys Rev Lett*, 2008, **100**.
70. H. Lu, B. Isralewitz, A. Krammer, V. Vogel and K. Schulten, *Biophys. J.*, 1998, **75**, 662-671.
71. J. R. Forman, Z. T. Yew, S. Qamar, R. N. Sandford, E. Paci and J. Clarke, *Structure*, 2009, **17**, 1582-1590.
72. M. Sikora and M. Cieplak, *Proteins: Struct., Funct., Bioinf.*, 2011, **79**, 1786-1799.
73. M. Sikora, J. I. Sulkowska, B. S. Witkowski and M. Cieplak, *Nucleic Acids Res.*, 2011, **39**, D443-450.
74. A. Valbuena, J. Oroz, R. Hervás, A. M. Vera, D. Rodríguez, M. Menéndez, J. I. Sulkowska, M. Cieplak and M. Carrión-Vázquez, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 13791-13796.
75. S. P. Ng, K. S. Billings, T. Ohashi, M. D. Allen, R. B. Best, L. G. Randles, H. P. Erickson and J. Clarke, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 9633-9637.
76. D. L. Guzmán, A. Randall, P. Baldi and Z. Guan, *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 1989-1994.
77. M. M. Balamurali, D. Sharma, A. Chang, D. Khor, R. Chu and H. Li, *Protein Sci.*, 2008, **17**, 1815-1826.
78. T. I. García, A. F. Oberhauser and W. Braun, *Proteins*, 2009, **75**, 706-718.
79. D. Sharma, O. Perisic, Q. Peng, Y. Cao, C. Lam, H. Lu and H. B. Li, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 9278-9283.
80. N. Bhasin, P. Carl, S. Harper, G. Feng, H. Lu, D. W. Speicher and D. E. Discher, *J. Biol. Chem.*, 2004, **279**, 45865-45874.
81. H. Li and J. M. Fernandez, *Journal of Molecular Biology*, 2003, **334**, 75-86.
82. M. S. Li, *Biophys. J.*, 2007, **93**, 2644-2654.
83. A. Galera-Prat, A. Gomez-Sicilia, A. F. Oberhauser, M. Cieplak and M. Carrion-Vazquez, *Curr. Opin. Struct. Biol.*, 2010, **20**, 63-69.
84. E. Eyal and I. Bahar, *Biophys J*, 2008, **94**, 3424-3435.
85. A. Srivastava and R. Granek, *Phys Rev Lett*, 2013, **110**.
86. M. Schlierf and M. Rief, *J. Mol. Biol.*, 2005, **354**, 497-503.
87. Y. Taniguchi, D. J. Brockwell and M. Kawakami, *Biophys. J.*, 2008, **95**, 5296-5305.
88. K. W. Plaxco, K. T. Simons and D. Baker, *J. Mol. Biol.*, 1998, **277**, 985-994.
89. Z. B. Guan, D. L. Guzman, A. Randall and P. Baldi, *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 1989-1994.
90. M. Sikora, J. I. Sulkowska and M. Cieplak, *PLoS computational biology*, 2009, **5**, e1000547-e1000547.
91. P. Zheng and H. Li, *J. Am. Chem. Soc.*, 2011, **133**, 6791-6798.
92. C.-L. Chyan, F.-C. Lin, H. Peng, J.-M. Yuan, C.-H. Chang, S.-H. Lin and G. Yang, *Biophys. J.*, 2004, **87**, 3995-4006.
93. A. F. Oberhauser, C. Badilla-Fernandez, M. Carrion-Vazquez and J. M. Fernandez, *J. Mol. Biol.*, 2002, **319**, 433-447.
94. M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke and J. M. Fernandez, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 3694-3699.
95. P. Zheng and H. Li, *Biophys. J.*, 2011, **101**, 1467-1473.
96. Y. Cao and H. Li, *Nat. Mater.*, 2007, **6**, 109-114.
97. D. J. Brockwell, G. S. Beddard, E. Paci, D. K. West, P. D. Olmsted, D. A. Smith and S. E. Radford, *Biophys. J.*, 2005, **89**, 506-519.
98. C. He, G. Z. Genchev, H. Lu and H. Li, *J. Am. Chem. Soc.*, 2012, **134**, 10428-10435.
99. H. Li and J. M. Fernandez, *J. Mol. Biol.*, 2003, **334**, 75-86.
100. K. L. Fuson, L. Ma, R. B. Sutton and A. F. Oberhauser, *Biophys J*, 2009, **96**, 1083-1090.
101. R. B. Best, B. Li, A. Steward, V. Daggett and J. Clarke, *Biophys. J.*, 2001, **81**, 2344-2356.
102. G. Arad-Haase, S. G. Chuartzman, S. Dagan, R. Nevo, M. Kouza, B. K. Mai, H. T. Nguyen, M. S. Li and Z. Reich, *Biophys. J.*, 2010, **99**, 238-247.
103. X. Gao, M. Qin, P. Yin, J. Liang, J. Wang, Y. Cao and W. Wang, *Biophys. J.*, 2012, **102**, 2149-2157.
104. M. Rief, J. Pascual, M. Saraste and H. E. Gaub, *J. Mol. Biol.*, 1999, **286**, 553-561.
105. C. Cecconi, E. A. Shank, C. Bustamante and S. Marqusee, *Science*, 2005, **309**, 2057-2060.
106. J. P. Junker, F. Ziegler and M. Rief, *Science*, 2009, **323**, 633-637.
107. H. Dietz and M. Rief, *P Natl Acad Sci USA*, 2004, **101**, 16192-16197.
108. 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, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 139-144.

Table 1. Proteins studied experimentally by force spectroscopy and determined Δx_U , sorted by decreasing unfolding force at a pulling speed of 600 nm/s. Values for the unfolding force at 600 nm/s have been interpolated where necessary. Δx_U values for C2A, C2B and barnase were determined by the correlation between F_U and Δx_U in this survey and are given in brackets. An asterisk (*) mark clamp motifs that have been assigned by the authors of this article. The given root mean squared error (RMSE) is based on the differences between interpolated and predicted unfolding forces within a pulling speed range of 100 to 1000 nm/s. RMSE containing cells are shaded according to their respective percentile of overall distribution of RMSE values showing the 75th percentile in red, the 25th to 75th percentile in yellow and the lower 25th percentile in green. A low RMSE value indicate that the protein behaves close to the average observed behaviour across the experimental data set. Higher deviations from the expected average behaviour are found for I27, I27mut, TnFNIII, I1 that are less force sensitive than predicted and Fe-pfRD, Zn-pfRD, C2B that are more sensitive to an applied force. For latter two rubredoxins, this may reflect that ferric- and zinc-thiolate bonds instead of H-bonds primarily mediate the mechanical strength of these proteins.

Protein name	pdb	F at 600nm/s (pN)	Δx_U (nm)	k_U (ms ⁻¹)	F x x_U (pN m)	N (aa)	Clamp motif	SCOP class	pulling speed range (nm/s)	RMSE (pN)	reference
Fe-pfRD	1BRF	230	0.14	150	32	53	-	small proteins (all beta)	100-4000	8.1	Zheng & Li, 2011a ⁹¹
Ubq	1UBQ	227	0.225	-	51	76	SS	alpha and beta	50-11000	3.6	Chyan et al., 2004 ⁹²
Ubq (N-C)	1UBQ	227	0.25	-	57	76	SS	alpha and beta	40-1110	2.0	Carrion-Vazquez et al., 2004 ⁸
1FNIII	1OWW	224	0.17	4	38	98	SD2	all beta	60-6000	1.4	Oberhauser et al., 2002 ⁹³
I27	1TIT	217	0.25	0.33	54	89	S	all beta	10-8000	5.7	Carrion-Vazquez et al., 1999 ⁹⁴
Zn-pfRD	1ZRP	198	0.14	100	28	53	-	small proteins (all beta)	100-4000	7.7	Zheng & Li, 2011b ⁹⁵
Protein G	1PGA	190	0.17	39	32	56	SS	alpha and beta	10-5000	0.2	Cao & Li, 2007 ⁹⁶
I27mut	1TIT	176	0.28	2	49	89	S	all beta	100-2000	8.1	Hoffmann et al., 2013 ⁴³
Top7	1QYS	165	0.21	60	35	92	-	Designed Proteins (all beta)	40-4000	1.3	Sharma et al., 2007 ⁷⁹
Tn 3FNIII	1TEN	160	0.3	0.46	48	81	SD2	all beta	10-1110	5.9	Oberhauser et al., 1998 ¹¹
Protein L	1HZ6	151	0.22	50	33	67	SS	alpha and beta	40-4000	0.8	Brockwell et al., 2005 ⁹⁷
AVF3-109	2J6B	115	0.24	1800	28	109	-	alpha and beta	50-4000	2.7	He et al., 2012 ⁹⁸
I1	1G1C	114	0.35	5	40	97	SD2	all beta	20-4000	4.1	Li & Fernandez, 2003 ⁹⁹
13FNIII	1FNH	98	0.34	22	33	90	SD2	all beta	50-5000	0.4	Oberhauser et al., 2002 ⁹³
C2B	1TJX	97	(0.41)	-	(39.4)	159	SD2 ⁺	all beta	50-5000	4.5	Fuson et al., 2009 ¹⁰⁰
10FNIII	1FNF	81	0.38	20	31	93	SD2	all beta	60-6000	0.5	Oberhauser et al., 2002 ⁹³
TmCspB	1G6P	80	0.49	12	39	66	SD2	all beta	100-2000	2.4	Hoffmann et al., 2013 ⁴³
Barnase	1BNR	68	(0.58)	0.0234	(39.4)	110	Z	alpha and beta	100-5000	3.3	Best et al., 2001 ¹⁰¹
C2A	2R83	55	(0.72)	-	(39.4)	124	SD2 ⁺	all beta	50-5000	3.6	Fuson et al., 2009 ¹⁰⁰
AcP	1APS	53	0.6	30	32	98	SS*	alpha and beta	100-10000	1.7	Arad-Haase et al., 2010 ¹⁰²
ddFLN4	1KSR	50	0.5	350	25	100	Z	all beta	200-4000	1.8	Schwaiger et al., 2004 ⁴⁸ , Schlierf & Rief, 2005 ⁸⁶
PAS-B	1X00	33	2	0.03	65	119	Z*	alpha and beta*	50-3600	2.4	Gao et al., 2012 ¹⁰³
Spectrin	1AJ3	32	1.7	-	54	110	Za*	all alpha	80-800	1.6	Rief et al., 1999 ¹⁰⁴
RNase H	1RNH	20	2	0.3	41	155	D	alpha and beta	10-1000	1.3	Cecconi et al., 2005 ¹⁰⁵
Cam DomC	1CFC	18	2	20	36	70	-	all alpha	10-250	0.9	Junker et al., 2009 ¹⁰⁶

15

20

25

Table 2. Proteins studied experimentally by force spectroscopy where only the unfolding force at one pulling speed has been reported. Values in brackets are estimations based on the found correlations reported here. * Δx_U derived from Monte Carlo simulations but no speed dependence of unfolding force given; †mechanical clamp motif self-assigned; [§] Δx_U derived from $\Delta\Delta G = RT \ln(k_1/k_2) = F\Delta x_U$ relation

Protein name	pdb	F_U at 600 nm/s (pN)	Δx_U (nm)	k_U (ms^{-1})	$F \times x_U$ (pN m)	N (aa)	Clamp motif	SCOP class	experimental pulling speed (nm/s)	reference
scaffoldin c7A	1AOH	(510)	0.11*	0.3	56	147	SD1	all beta	400	Valbuena <i>et al.</i> , 2009 ⁷⁴
scaffoldin c1C	1G1K	(452)	0.133*	0.2	60	143	SD1	all beta	400	Valbuena <i>et al.</i> , 2009 ⁷⁴
scaffoldin c2A	1ANU	(228)	0.17*	10	39	138	SD1	all beta	400	Valbuena <i>et al.</i> , 2009 ⁷⁴
GFP	1B9C	(116)	0.28*	-	33	238	SD1	all beta	300	Dietz & Rief <i>et al.</i> , 2004 ¹⁰⁷
T4 lysozyme	1B6I	(60)	0.81 [§]	0.1	49	164	Za	alpha and beta	1000	Yang <i>et al.</i> , 2000 ¹⁰⁸
VCAM1	1VCS	(37)	1.6*	$\frac{0.000}{1}$	60	102	Z	all alpha	1000	Bhasin <i>et al.</i> , 2004 ⁸⁰