FEATURE ARTICLE

ChemComm
www.rsc.org/chemcomm

What can atomic force microscopy tell us about protein folding?

Robert B. Best and Jane Clarke*
University of Cambridge Department of Chemistry, MRC Centre for Protein Engineering, Lensfield Road, Cambridge, UK CB2 1EW

Received (in Cambridge, UK) 9th September 2001, Accepted 11th November 2001
First published as an Advance Article on the web 13th December 2001

Force spectroscopy has emerged as a new tool to study protein folding, in which force replaces the chemical denaturant used in traditional folding experiments. This new technique complements older methods and allows a range of new questions to be investigated. What sort of protein is able to resist mechanical unfolding, and to what extent is mechanical stability dictated by fold or function? What is the effect of force on the unfolding energy surface? Do proteins unfold by the same pathway in mechanical and chemical denaturation experiments? Answers to these are starting to emerge based on a combination of experimental and computational approaches. We present some of the forced unfolding experiments to date and simple methods for characterizing the unfolding potential from the results. Several studies have also begun a more fine-grained description of mechanical unfolding, for example by invoking intermediates to explain features seen in unfolding traces and by using mutagenesis to try to localize the origin of mechanical stability. We propose further experimental approaches to this goal using the protein engineering method to characterize transition states, similar to those used in conventional folding experiments. However, it is likely that a high-resolution picture of mechanical unfolding will only emerge through a combined interpretation of careful experimental work and computer simulation.

Introduction

Unlike simple chemical reactions, the energy surface describing the process by which proteins fold into their unique[1] native structure is highly multidimensional due to the immense number of degrees of freedom of the polypeptide chain. If the protein were to search randomly through all potential conformations for the global minimum (native state) it would take an unrealistically long time to fold, yet it is well known from experiment that this process takes from microseconds to seconds.[2,3] This paradox, first stated by Levinthal,[4] is often explained by considering the general shape of the energy surface, or energy landscape.[5–7] In order to fold efficiently, a protein must have a 'funnel-shaped' energy landscape leading down to the native state, without any significant kinetic traps; thus, far from randomly searching through equally probable conformations, the protein is guided by an energetic bias. Much work has been done to understand the nature of this energy surface by considering analytical[8,9] and simplified lattice models[10,11] as well as all-atom molecular dynamics simulations.[12–14] Experimentally, folding is most often studied on a bulk solution sample using chemical denaturants such as urea or acid to accelerate the rate of unfolding.

Recently, experiments using the atomic force microscope (AFM)[15] in force mode and related force probe methods[16] (together known as *force spectroscopy*) have provided a novel method of studying protein unfolding, with force playing the role of 'denaturant',[17–21] which can potentially give new insights into the protein folding process and energy landscapes. These techniques allow the unfolding of individual domains in a multi-modular protein to be monitored as it is stretched. The AFM experiment differs in a number of important ways from the more conventional bulk solution experiments. Firstly, it is a single-molecule experiment, so that it is potentially possible to observe differences in the nature of unfolding events from molecule to molecule, one of the few methods of doing this (fluorescence energy transfer and fluorescence quenching have been used to study single molecule protein folding in chemical denaturant[22,23]). Secondly, the effect of force can be mapped onto a well-defined reaction coordinate, namely the N–C terminal extension, in contrast with denaturant-induced unfolding. For this reason, it is a good technique for comparison with computer simulations of unfolding, which similarly use single molecules. Finally, perhaps most importantly, it offers the possibility of probing alternative regions of the energy landscape to those explored in experiments with chemical denaturant.

In this article we will try to provide a framework for investigating protein folding using AFM, although much of the discussion applies to other forms of force spectroscopy as well. We describe methods for interpreting AFM experiments on proteins in terms of energy landscapes, what parameters can be extracted to describe the unfolding kinetics and how reliable these are. We address some questions that are raised by the new experiments. Which proteins are able to resist force well? Clearly all proteins are different, but is it the overall fold or physiological function that primarily explains the origin of mechanical strength? Do proteins unfold by the same 'pathway' under force as they do physiologically, or in studies with chemical denaturant? Finally, insights have been given in the literature on unfolding pathways and intermediates in force-induced unfolding. We propose a method for studying in detail

Robert Best was born in Cape Town in 1976. He obtained his BSc(Hons) degree in Chemistry from the University of Cape Town in 1998, followed by an MSc in computational studies of carbohydrates, supervised by Dr Kevin Naidoo and Professor Graham Jackson. He is currently studying protein folding for his PhD at the University of Cambridge under the supervision of Dr Jane Clarke.

Jane Clarke was born in London in 1950. After achieving a degree in Biochemistry at York in 1971 she went into teaching and taught Science (Biology/Chemistry) in comprehensive schools in London and elsewhere. In 1990 she 'went back to school' and started a PhD with Professor Alan Fersht in the MRC Centre for Protein Engineering in Cambridge. She has stayed in Cambridge since then and has just been awarded a five-year Wellcome Trust Senior Research Fellowship, which she holds in the Chemistry Department at Cambridge University. Her research is centred around the study of protein folding, combining protein engineering, biophysical chemistry, structural biology (mainly NMR) and simulation to address the challenge posed by the genome projects: How far can we predict the structure and properties of a protein from its sequence?

DOI: 10.1039/b108159b

the pathway of forced unfolding, analogous to conventional Φ-value analysis.

Pulling proteins by AFM

The force-mode AFM experiment

A schematic illustration of a typical force-mode AFM experiment for stretching proteins is given in Fig. 1(a). A multi-modular protein is tethered at one end to a microscope slide (a frequently used method is the strong adsorption of a pair of C-terminal cysteines to a gold surface[24]) and at the other to a silicon nitride cantilever (usually by non-specific adhesion). The cantilever is pressed against the microscope stage for a short period of time in order to allow adhesion. Movement of either the headpiece or the microscope stage causes an extension of the protein and the force exerted can be measured by the deflection of the cantilever, if indeed a protein was picked up. Calculation of the force depends on knowing the spring constant of the cantilever, which can be calibrated from its free thermal fluctuations in water.[25]

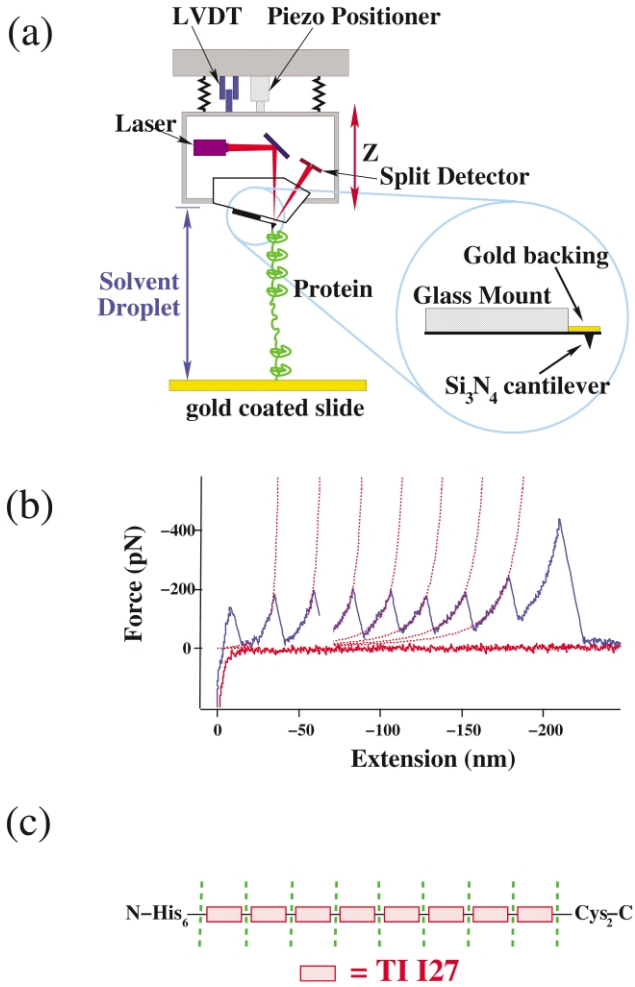

Fig. 1 The force-mode AFM technique. (a) Diagram of an atomic force microscope: the protein is attached by gold–sulfur bonds at its C-terminus to a gold-coated microscope slide and by non-specific adhesion to the tip of the AFM cantilever; the protein is completely covered by a layer of buffer, several millimeters from the solvent surface. The cantilever itself consists of a thin lever of silicon nitride, the back of which is coated with a reflective layer of gold. Cantilever deflection is measured by the deflection of a laser light reflected off the back of the cantilever. (b) Typical force-extension trace for pulling an octameric TI I27 poly-protein. Note that the extension is that of the protein itself, obtained by subtracting the cantilever deflection from the position of the cantilever base set by the piezo. The dotted line indicates fit of the worm-like chain model to the back of each peak. (c) Example of a gene construct which has been used to clone multiple copies of the same protein: the positions of restriction sites used for its assembly are shown by dotted lines (A. Steward, J. Clarke *unpublished*).

Fig. 1(b) shows a typical saw-tooth force-extension trace that was obtained from a pull at constant pulling speed, in this case for a designed poly-protein of eight titin I27 (TI I27) modules (titin is a 3 MDa muscle protein that is responsible for exerting a passive restoring force on the sarcomere). An initial peak is usually observed as the tip lifts off the surface (zero extension in Fig. 1(b)), even when no protein has been 'picked up', and is attributed to the interaction of the tip with the surface or proteins on the surface. It may also include the unfolding of the first module, but this cannot be determined with certainty. Each peak following this has a curved, sloped back, interpreted as the soft elastic extension of already-unfolded protein. This is followed by a sudden drop in the force as each domain unfolds in turn, causing an increase in the total protein length. Detachment of the protein from the tip is responsible for the final peak, which is usually much larger than the others. We can see that only seven out of eight modules unfolded in the example trace, presumably because the eighth was involved in binding to the tip. Refolding experiments are not possible under tension as the denatured (extended) state is stabilized too much by force,[17] but stretch–relax cycle refolding experiments have been done, in which the AFM tip is lowered after an initial pull and then retracted again after waiting a certain time. Refolding rate constants (at zero force) can be calculated from the number of peaks remaining after different waiting times.[26a] These refolding experiments give basically the same information as obtained in conventional refolding experiments with chemical denaturant, although they have allowed rare misfolding events to be observed;[26b] they will not be discussed here.

Multi-modular protein constructs

A multi-modular protein is preferred for a number of reasons. Most obviously, because we are relying on non-specific attachment of the protein to the AFM tip, and the N-terminal attachment point can be anywhere along the length of the protein, having several domains in tandem increases the chance of picking up a reasonable length of protein. Due to the length of the protein, most of the modules will unfold when the tip is reasonably far from the surface (>50 nm), so that any interactions of the tip with the gold surface or with other proteins on the surface will be negligible. In addition, the first peak arising from the detachment of the tip from the surface can sometimes be quite large (possibly due to additional proteins being attached), and so the force-extension response of a single protein module would likely be lost in the noise. Finally, having a number of modules increases the amount of data that can be obtained from each trace and the correspondence of the number of force peaks with the number of protein modules and their overall similarity gives us confidence in the interpretation that each peak is caused by the unfolding of a single module.

Initially, recombinant constructs consisting of segments of native multimodular proteins were used, in which the modules are all different, although they usually have the same overall topology.[17,18] It is not possible to assign the peaks in the trace to particular modules for such proteins, apart from making the trivial observation that the 'weakest' modules unfold first. Unambiguous interpretation necessitates a poly-protein consisting of only one type of module. Several designs for the construction of such a protein using molecular biology techniques have been proposed;[26a,27] an example is given Fig. 1(c) in which unique restriction sites were used to 'paste' together several domains.[28] An advantage of the use of unique restriction sites is that other modules can later be inserted into the poly-protein.

Interpretation of force data

A set of AFM data contains only a few 'ideal' force-extension curves such as that shown in Fig. 1(b). Many traces are less easy

to interpret, often because of the attachment of more than one protein. Such events are typified by higher than expected forces and peaks which are closer together than usual, due to a partial offset in the points of attachment of the proteins; sometimes the second protein becomes detached, after which a normal trace is observed (Fig. 2 shows an example of this). To avoid ambiguity,

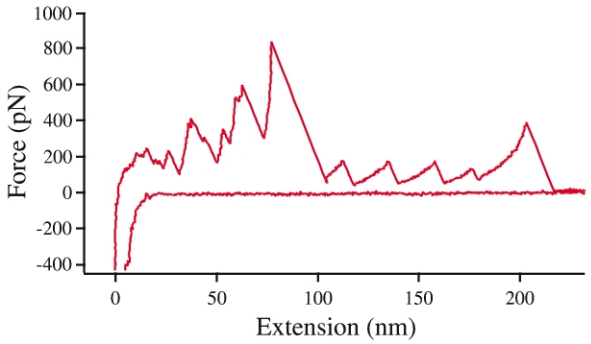

Fig. 2 Sample force–extension trace believed to arise from two proteins being pulled at once.

all of these traces are excluded from further analysis. Also beyond our control is the position along the poly-protein at which tip attachment takes place. The result of this is that the number of peaks in a trace can vary between one peak and the full number present in the construct. Since it is not improbable that a trace with only one or two peaks is caused by the detachment of weakly bound protein molecules, we use only traces with at least three peaks.

Can the force data from each unfolding event in a pull be pooled, or must the data from the first peaks, second peaks and so on be separated? In principle, the position of the peak in the force trace (first, second, *etc*.) is important: the most probable unfolding force is related to the number of modules remaining folded, and to the rate at which force is applied, both of which are changing as more modules become unfolded. At present, however, the data from all peaks are generally pooled, in order to maximise the amount of data used—this is justified by the small dependence of force on the order of the unfolding event compared with the large spread of the data (J. L. Toca-Herrera, J. Clarke, *unpublished data*). Provided there is sufficient data, it is more correct to separate it by unfolding order; furthermore, the effect of peak order on unfolding forces could potentially provide additional information.

Determining protein length

An estimate of the unfolded length of the protein can be made by fitting it to the expected force–extension profile of each peak for a simple polymer model. The model most frequently used is the worm-like chain, an approximation to which is given in eqn. (1).[17]

$$F(x) = \frac{k_B T}{p}\left\{\frac{1}{4(1-x/L)^2} + \frac{1}{4} - \frac{x}{L}\right\} \qquad (1)$$

This model characterizes the polymer in terms of its total length or contour length, L, and its persistence length, p, which is a measure of flexibility (the correlation length of a tangent to the backbone); F is the force at polymer extension x, while k_B and T are the Boltzmann constant and temperature respectively. Even though this model does not account for the complexity of the peptide chain, it turns out to fit the data well. If the value of the persistence length is fixed, the fit of the worm-like chain to the back of each peak (dotted lines in Fig. 1 (b)) gives an estimate of the change in contour length between adjacent peaks; in the case of titin I27 this is approximately 295 Å which is equivalent to the difference in length between a folded and unfolded module. The consistency of this difference supports the interpretation of each peak as the unfolding of a single module. Moreover, the change in fitted contour length, ΔL, can also be used to identify cases in which two proteins have been picked up.

Distribution of unfolding force

Due to the 'single-molecule' nature of the experiment, there is an inherent thermal distribution of unfolding forces obtained from AFM traces. The standard deviation of this spread is approximately $\sigma_f \approx k_B T/x_u$, where x_u is a parameter describing the width of the unfolding potential (see below).[29] On top of this there are many sources of error in current instruments. Most of the errors can be attributed to the cantilever, whose mechanical properties govern the sensitivity and accuracy of the measurements. Soft cantilevers are more sensitive to changes in force, but suffer from thermal fluctuations in position which affect the accuracy of force measurements; stiff cantilevers, while reducing thermal noise, increase the size of fluctuations transmitted to the protein by the cantilever (although these ought to be attenuated by the soft elastic linkage of the unfolded protein). However, this effect can, in principle, be reduced by making sufficient measurements. Systematic errors in the measured force can arise from hydrodynamic effects: the viscous drag on the cantilever creates a speed-dependent force, which is difficult to account for, as it should depend on the distance of the cantilever from the surface. There is evidence for it in the slight offset between approach and retraction traces at high pulling speeds: the exact magnitude of the effect will depend on the cantilever and should be reduced by the use of the very small cantilevers currently being developed.[30,31] Errors may also arise from the calibration of different cantilevers: as many different cantilevers are usually used in the collection of a data set, any calibration errors will add to the data noise. An estimate of the calibration error—as well as other systematic day-to-day differences—can be made from the scatter of the mean unfolding force per tip—typically the standard deviation of the mean data collected by different cantilevers is 10–20 pN.

In addition to the cantilever, there is drift in the vertical displacement of the cantilever. This is partly due to drift in the piezo positioner, but this can be corrected for by measuring the piezo displacement by an alternative technique. A second source of error comes from thermal expansion and contraction of the entire headpiece. In order to make quantitative measurements all these effects need to be considered.

Unfolding forces depend on pulling speed

Although we speak of 'mechanical stability', we need to distinguish whether this is a kinetic or thermodynamic property of the protein being studied. Fig. 3 shows the mean unfolding force for TI I27 over a range of pulling speeds. The observed linear increase of unfolding force with pulling speed is a general property of all unfolding studies of proteins by AFM to date. If the measured force were an equilibrium property of the protein, it should depend only on the state variables describing the system, and therefore only on end-to-end extension (the only such variable which is changed in these experiments). The fact that it depends also on pulling speed demonstrates that this is a kinetic phenomenon. If we assume that increasing force reduces the energy barrier to unfolding monotonically, the probability of unfolding is increased with extension; however, the likelihood of thermal escape at each extension is also dependent on the time spent at that force, hence the dependence on pulling speed.[32,33] This holds the key to extracting kinetic information from these experiments; it is important therefore to consider the pulling speed when comparing unfolding forces.

Which proteins are able to resist force?

Most of the proteins that have been used for forced unfolding measurements have been intentionally selected for their func-

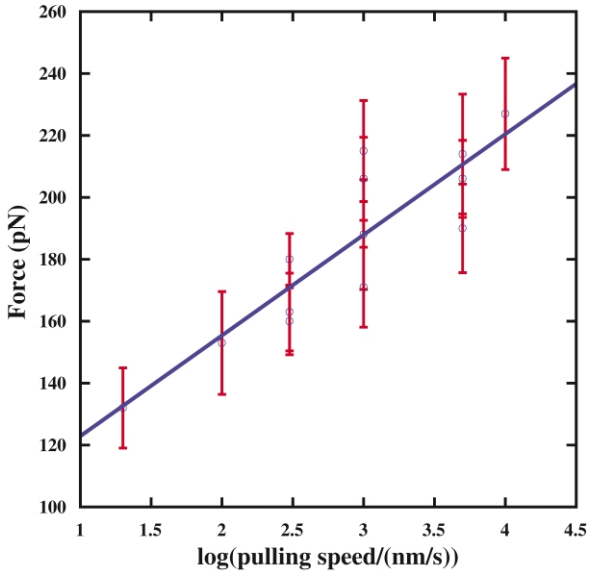

Fig. 3 Dependence of unfolding forces of titin I27 on pulling speed. Each point is the average of a set of unfolding force data collected at that pulling speed with the same cantilever; the large scatter in the data is due partly to the thermal noise in the experiment and partly to the inherently broad distribution of single-molecule data.

tional requirement to resist force. They are, for example, muscle proteins (immunoglobulin domains from the I band[17,26a] and the fnIII domains from both I and A bands[18] of titin), cytoskeletal proteins (spectrin[19,20]) or extracellular matrix proteins involved in cell adhesion (tenascin,[34] fibronectin[35] and Ig-CAM[36]). Even amongst these proteins that require mechanical strength there is a considerable range in unfolding forces, from spectrin (25–35 pN) to titin I28 (250-280 pN). In addition to these, the forced unfolding of the enzymes T4 lysozyme[21] and barnase[28] has been reported, both of which show low unfolding forces. Can these results be rationalized in terms of the overall fold and topology of the domains and their evolutionary (functional) requirements? We have chosen here to focus on the examples shown in Fig. 4: namely titin I27, the third fnIII domain of human tenascin (TNfn3), α-spectrin, barnase and T4 lysozyme, since these have been well characterized both by solution unfolding and by AFM.

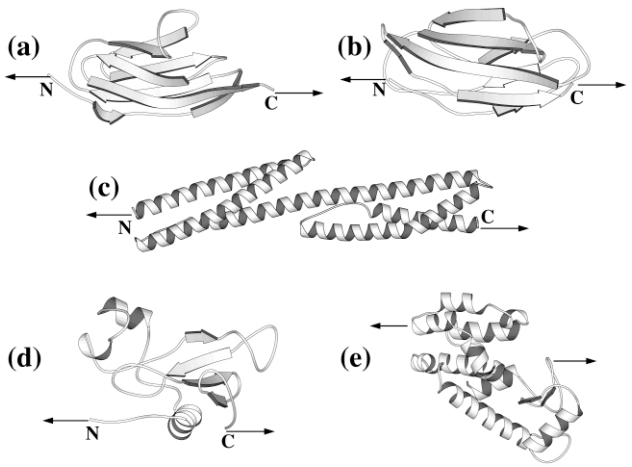

Fig. 4 Representative examples of proteins studied by AFM. (a) Titin I27, (b) third fnIII domain of human tenascin, (c) two domains from α-spectrin, (d) T4 lysozyme and (e) barnase. The N and C termini are indicated for all the proteins pulled from their ends and the location and direction of the pulling forces are shown by arrows.

Immunoglobulin-like β-sandwich domains

The immunoglobulin-like fold describes a large set of small protein domains with a common overall structure: 7 strands arranged in two facing β-sheets with Greek-key topology ('β-sandwich'). They are usually found in modular proteins, which consist of a tandem array of independently folding domains. The immunoglobulin (Ig) and fibronectin type III (fnIII) superfamilies are groups of proteins that belong to this fold and several examples of each have been studied by AFM. They are ideal for exploring the relationship of mechanical strength to structure and function, providing examples of variation in both properties; additionally, they can be used to test the importance of domain–domain interactions in mechanical stability since they naturally occur in multi-modular arrays. Titin I27 and the third fnIII domain from human tenascin (TNfn3) are given as examples of each superfamily in Fig. 4 (a) and (b) respectively.

The immunoglobulin domains from the I band of titin are the proteins most studied by AFM; their unfolding forces range from 150 to 300 pN at a pulling speed of 1 μm s^{-1} in full length native titin.[17] The actual variation may be less marked, as the early peaks in very long traces such as these will be inherently lower (there are more folded domains and hence a higher chance of one unfolding). More certain are the experiments comparing modules I27 and I28 in short designed constructs:[27] at a speed of 600 nm s^{-1}, the average unfolding force of I27 is 211 pN and that of I28 306 pN. The structure of I28 is unknown, but both domains are predicted to have essentially the same structure (in terms of topology and backbone hydrogen bonding) despite a fairly low sequence identity, suggesting that the side-chain interactions conferred by the sequence can significantly modulate mechanical stability.

The superfamily of fibronectin type III (fnIII) domains is one of the most abundant in the genomic database (2814 sequences in the most recent version of Pfam[37]), and its members are extremely diverse in function. The extracellular matrix proteins fibronectin and tenascin, involved in cell adhesion, both contain fnIII domains. AFM experiments on these proteins showed significantly lower forces than for the unfolding of titin I27 and I28: fnIII modules in native fibronectin unfold at between 90 and 150 pN at an extension speed of 1 μm s^{-1}[35] and those in native tenascin at an average of 137 pN (0.2–0.6 μm s^{-1})[34] and 113 pN (0.5 μm s^{-1}),[18] with a small dependence on pulling speed. This suggests that these proteins may have a lower functional requirement for mechanical strength than the titin Ig domains. Does the difference in mechanical strength mean that the Ig fold has inherently higher mechanical stability than the fnIII fold? The occurrence of fnIII domains in titin is able to provide an answer: sections of fnIII-rich titin from the constitutive region of the I band and from the A band have high average unfolding forces (200 pN and 180 pN respectively at a speed of 0.5 μm s^{-1}),[18] comparable to those of some Ig domains. Thus the mechanical strength of both the fnIII and Ig scaffolds can vary significantly with sequence details, permitting the adoption of different functions; however, it seems that in general the Ig topology can tolerate higher mechanical stress than the fnIII topology.

An interesting future possibility may be the investigation of inter-domain interactions (between adjacent modules), which have been suggested, but not satisfactorily confirmed, for several of these proteins. For instance, initial unfolding of modules I27 and I28 by AFM appeared to show that I28 was mechanically stabilized by the presence of I27;[27] however, bulk solution unfolding and stability experiments now suggest that this was an artifact caused by conservative selection of domain boundaries.[38] A second example is the A band of titin, which is rich in fnIII domains, some of which are separated by a linker of only two amino acids. The conserved residues of the A band fnIII domains which are specific to titin (as opposed to all fnIII proteins) consists of a patch on one face of each module

(thought to be associated with the other A band proteins) and a patch at the ends of the modules immediately adjacent to their neighbours, suggesting a possible role in domain–domain interaction.[39,40] However, there is no evidence of inter-domain interactions on the bulk unfolding kinetics and stability of the A band fnIII domains.[41] Thus far, conclusive evidence from AFM of inter-domain interactions between modular β-sandwich domains remains elusive. Nonetheless, not many cases have yet been systematically studied, and there are highly suggestive sequence and structural leads that need further investigation.

α-Helical domains: spectrin

In contrast to the Ig-like β-sandwich domains, the cytoskeletal protein spectrin has an all-α structure, composed of repeats of a three helix bundle in which adjacent domains share a continuous helix (Fig. 4 (c) shows two such domains). In the cell it forms an αβ heterodimer which associates head-to-head to form an $\alpha_2\beta_2$ tetramer. Force peaks in spectrin traces are much lower than those in titin, being approximately 30 pN at a pulling speed of 0.5 μm s^{-1},[20] a difference that is significantly greater than the variation within the β-sandwich fold. Here, it seems probable that the differences in force are due to the mechanical weakness of this kind of fold, which has been attributed to the lack of stabilization against force by hydrogen bonds in the secondary structure.[20]

Non-mechanical proteins

There is clearly a large variation in mechanical strength for proteins which need to withstand force, but will non-mechanical proteins resist force as well? The protein folding models barnase and T4 lysozyme have been used to address this question, since they have been well characterized by bulk solution studies.[42–44] Since both proteins are enzymes and not naturally multimeric, strategies were devised to link them together. T4 lysozyme (Fig. 4 (d)) was ingeniously polymerized in the solid state by disulfide cross-linking of engineered cysteines.[21] The resulting oligomers showed regularly spaced force peaks with an average unfolding force of 64 pN at 1 μm s^{-1}, which could each be fit to the worm-like chain model with constant contour length increments between them. A structural reason for this lower resistance to force could be that the protein is all helical, like spectrin, and lacks the mechanical stabilization by hydrogen bonds between β-strands in the β-sandwich domains.

Barnase (Fig. 4 (e)) was spliced into the existing titin I27 multimeric gene construct shown in Fig. 1(c), replacing modules 2, 4 and 6.[28] This has a number of advantages: because the modules in the multimer are separated by spacers of several residues, interactions between neighboring domains are less likely than for the disulfide-linked lysozyme. In addition, the remaining I27 modules are able to provide an internal reference for comparison of unfolding forces, and in practice aid in the intracellular expression of the protein. Barnase also unfolds at low forces (70 pN at 0.3 μm s^{-1}), clearly distinct from the I27 modules, but with irregularly spaced peaks. The low resistance to force may be explained by the fact that the main constraints on its structure are tertiary; molecular dynamics simulation results show that the main force peak corresponds to the disruption of the hydrophobic core.

The effect of force on the unfolding energy landscape

The initial use of the AFM for the mechanical unfolding of proteins gave results of primarily qualitative value; however, as the force of unfolding is determined by the unfolding kinetics, an interpretation in terms of the unfolding energy barrier should be possible. From this, a picture of the effect of force on the energy landscape can be constructed.

How does force affect energy landscapes?

Unlike denaturant, which has a uniform effect on the protein structure, causing a stabilization of more expanded, solvent-exposed structures, force is directional. Large forces cause proteins to adopt high energy extended structures that would never usually be accessed, even by 'unfolded' protein. The effect of force has been investigated using lattice model simulations, in which the residues of the protein are represented by connected beads on a lattice with simple contact potentials, and analogous off-lattice models, permitting a thorough statistical analysis of the (un)folding process. Simulations incorporating the effect of tension using a potential of mean force have shown three regimes of force: low force has little effect beyond polarizing the structure, intermediate forces give rise to exponential kinetics, the height of the unfolding barrier being lowered in approximate proportion to the force and high forces overwhelm the barrier, so that increasing the force further has little effect on kinetics.[45] The second, exponential, regime in which kinetics depends on force is evidently that sampled by AFM experiments. It is worth noting that the unfolded state of the experimental regime is very extended, compared with forces close to equilibrium, in which the unfolded state is only slightly more expanded than the native state. If end-to-end extension is used as the reaction coordinate, the barrier to unfolding is quite steep and its position on this coordinate does not vary greatly with force.[45,46]

For the purposes of interpreting AFM data, a simplified model of the landscape, which describes only the relative energies of the kinetically relevant states and their position on the reaction coordinate is necessary. A commonly used description is a two-state model specified by the energies of the native, denatured and transition states and their positions.[47] Fig. 5 shows a schematic diagram of the energy barrier. In the absence of force, the folding and unfolding reaction rates (k_f^o and k_u^o) are given using transition state theory by eqn. (2), in which ΔG_{D-TS} ΔG_{TS-N} are the activation energies for folding and unfolding respectively, k_B is the Boltzmann constant, T is absolute temperature , ν is the characteristic vibration frequency at the saddle point and κ the transmission coefficient.†

$$k_f^o = \kappa\nu \exp\left(\frac{\Delta G_{D-TS}}{k_B T}\right) \qquad k_u^o = \kappa\nu \exp\left(\frac{-\Delta G_{TS-N}}{k_B T}\right) \quad (2)$$

The effect of force is considered to add a term $-F{\cdot}x$ to the potential surface as a function of the distance x from the native state, a model first proposed by Bell.[49] If the distance along the unfolding reaction coordinate (assumed to be end-to-end extension) from the native to transition state is given by x_u and the distance from the transition state to the denatured state by x_f, then the energy barrier for unfolding will be lowered by $F{\cdot}x_u$ and that for refolding raised by $F{\cdot}x_f$. Implicit is the assumption that the position of the transition state does not change with increasing force. The rates for unfolding and refolding as a function of applied force are then as in eqns (3.1) and (3.2).

$$k_f(F) = \kappa\nu \exp\left(\frac{\Delta G_{D-TS} - F\cdot x_f}{k_B T}\right) = k_f^o \exp\left(\frac{-F\cdot x_f}{k_B T}\right) \quad (3.1)$$

$$k_u(F) = \kappa\nu \exp\left(\frac{-(\Delta G_{TS-N} - F\cdot x_u)}{k_B T}\right) = k_u^o \exp\left(\frac{F\cdot x_u}{k_B T}\right) \quad (3.2)$$

In practice, due to the large distance x_f between the unfolded state and transition state for forced unfolding of proteins, the probability of refolding events is negligible, and only unfolding rates play a part in the kinetics. Thus, only two parameters are needed to characterize the simplified unfolding energy landscape, namely x_u and k_u^o. The approximation for the acceleration of unfolding by force given in eqn. (3.2) is independently justified by its ability to reproduce the unfolding rates in lattice models using a biasing force.[46]

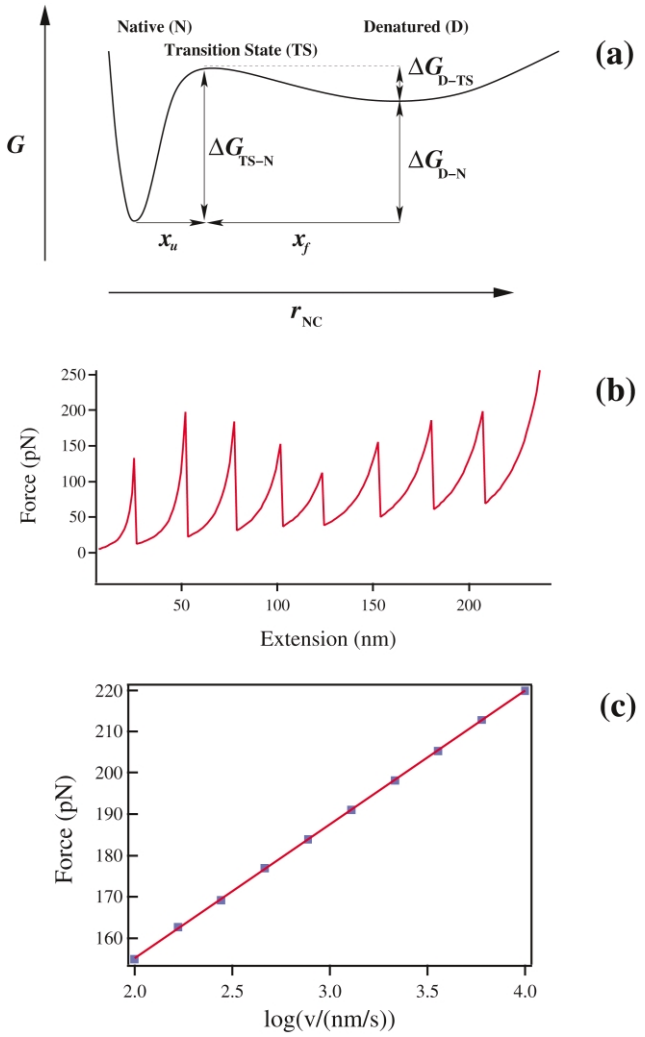

Fig. 5 (a) Two-state model for interpreting AFM experiments. Sample (b) force distribution and (c) pulling speed dependence obtained from Monte-Carlo simulations using the two-state model.

Determining kinetic parameters by Monte-Carlo simulation

The rate constants (or turnover times) given in eqn. (3) are unfortunately not directly determined in the AFM experiment in which the pulling speed and not the force is constant, and the force varies over a wide range, due to the soft elastic linkage with the unfolded protein (a constant force AFM has recently been developed which should help overcome these problems of interpretation).[50] However, one can invert the problem and use the above expressions for the rates under force in Monte-Carlo simulations to calculate unfolding forces at any given pulling speed (analytical approximations based on similar assumptions have also been proposed).[33] The simulations are performed by starting with N folded domains and increasing extension in increments of $v\Delta t$ at time intervals Δt, where v is the pulling speed. The force at each time t is determined by the elasticity of the unfolded protein (the tip is relatively rigid and can be neglected to a first approximation). Force is calculated from a worm-like chain model with extension $v\Delta t$, the contour length being known from the number of domains unfolded up to time t. Then the probability of a single domain independently unfolding during each time interval is $p_u = k_u(F)\Delta t$, and the probability of only one in n remaining folded domains unfolding is given by the binomial distribution $p^n{}_u(F) = np_u(1 - p_u)^{n-1}$. The time interval Δt is chosen sufficiently small that the chances of more than one domain unfolding are negligible, in which case p_u is small enough that $p^n{}_u(F) \approx np_u$. A random number between 0 and 1 is used to determine whether unfolding takes place, in which case the number of folded domains is reduced by one and the length of unfolded protein is increased by the length of one unfolded module. The algorithm allows complete force–extension traces to be generated, comparable to the experimental ones, and can reproduce the variation of unfolding force with pulling speed. The parameters k_u^o and x_u can be optimized by iteratively improving the match between experimental and simulated data. An example of the fit of unfolding forces to titin I27 data is given in Fig. 5 (a), for which parameters of 3.0×10^{-4} s^{-1} and 2.5 Å were determined for k_u^o and x_u respectively.[26a]

The parameters derived from the two-state model can explain the mechanical strength of the proteins which have been studied. Table 1 lists k_u^o and x_u for several of these proteins. It is apparent that the difference in mechanical strength may come not from the height of the unfolding barrier at zero force, but rather from the distance from the native state to the barrier; for instance, spectrin, which has a similar unperturbed unfolding rate k_u^o to titin I27, unfolds at much lower force because of its 'wider' unfolding potential. The effect of force on the transition state is therefore greater. It must be emphasized that the different values of x_u explain the differences in the effect of force on the energy barrier for proteins with similar k_u^o; since mechanical stability has nothing to do with the absolute difference in energy between native and denatured states, it is neither necessary nor valid to invoke such an argument to rationalize mechanical stability in terms of thermodynamic stability as is often done.[20,41]

Table 1 Two-state model parameters for titin,[26a] spectrin[20] and barnase.[28] Note that the spectrin parameters are fitted to data from a native construct in which each module is different and cannot be directly compared with unfolding rates measured in bulk solution. The titin and barnase data are from designed constructs with identical modules.

	Titin I27	Spectrin	Barnase
k_u^o(s^{-1})	3.0×10^{-4}	3.0×10^{-5}	6.7×10^{-2}
x_u (Å)	2.5	17	3.5

Can a precise physical meaning be given to the parameters from the two-state model? The 'unfolding distance' x_u is extremely short for many of the proteins when compared with the values derived from molecular dynamics simulation;[51] if it can be given a simple structural interpretation, it would most likely correspond to the difference in extension between the transition state and the polarized native state (at low force) rather than the native state itself. That the values of k_u^o determined for titin I27 and I28 are very similar to those determined in solution is intriguing,[26a,27] and suggests that the pathways of unfolding in bulk solution and under force, if not actually the same, at least share a similar barrier height for these proteins.

The large thermal spread in the data, and the various sources of experimental error mean that there is a corresponding uncertainty in the values of the parameters fitted by Monte-Carlo simulation. As yet, no published work has given a quantitative estimate of these errors apart from a graphical illustration of different fits: this needs to be addressed in the near future, especially for the purposes of comparing data from different mutants or for quantitative comparison with bulk measurements.

The interpretation of the two state model given here assumes a single transition state which is invariant under force, that is, the position of x_u does not change with pulling speed. This appears to be a valid assumption over the range of forces that can be accessed by force-mode AFM, as the dependence of the unfolding force on pulling speed is always linear: this is suggests a single transition state, since the slope of the curve is related to x_u.[52] However, the range of speeds or force loading rates sampled by the AFM is very narrow, so it is possible that alternative transition states are exposed at different pulling

speeds. Indeed, multiple transition states have been inferred for ligand unbinding studies carried out over a wide range of loading rates, as the force-loading rate curves have a succession of linear regimes with slightly different slope.[53] The transition state measured by AFM experiments may thus not be the same as that in solution: future studies extending protein unfolding experiments to slower loading rates need to be carried out to address this issue.

Mechanical unfolding pathways from experiment and simulation

By themselves, force-mode AFM experiments at a number of pulling speeds can give basic information on the unfolding energy landscape *via* the parameters of the two-state folding model: the width of the unfolding potential, x_u, and the unfolding rate at zero force, k_u^o. Of course one would like to go beyond this to a more fine-grained description of unfolding pathways, and to compare these pathways with those in solution. Doing this requires a combination of careful AFM experiments and computer simulation. Molecular dynamics simulation with a force bias has proved valuable in understanding mechanical unfolding pathways, just as it has for conventional folding experiments.[13,14]

Are force- and denaturant-induced unfolding pathways the same?

The mechanical unfolding rate of titin I27 of 3.3×10^{-4} s^{-1}, obtained by extrapolating to zero force, is very close (within error) to that determined in bulk solution by extrapolating to 0 M denaturant (4.9×10^{-4} s^{-1}).[26a] It is tempting to conclude that the unfolding pathways in the two experiments are the same. Do the similar rate constants in fact indicate an identical pathway, or is this a coincidence, and will the result extend to other proteins? This has been investigated in detail for both the RNase barnase and titin I27, using AFM experiments and molecular dynamics simulations.

Molecular dynamics simulations of titin I27 using thermal unfolding at high temperature and mechanical unfolding using a force bias[51] have been done to compare the different denaturant effects. These suggest that the thermal and mechanical pathways are quite different. In the mechanical unfolding, a single intermediate is observed at low force, in which the A strand becomes detached from the body of the protein, leading to an extension of approximately 8 Å. The unfolding force peak is due to the separation of the A' and G strands, following which there is catastrophic unfolding without further significant barriers. The same features have been seen in a number of studies using related methods.[54–56] However in simulations of thermal unfolding, this strand remains attached until late in the unfolding pathway, and the overall pattern of unfolding is different. Heat causes disruption of the β-sheet, formation of short helical segments and an expansion of the structure. Force causes the strands to be sequentially pulled from the body of the protein with the remaining structure relatively unaffected. The only common feature is that the same elements of structure unfold last. If the thermal unfolding pathway is similar to that found with chemical denaturant, as is often assumed,[13,14] then the denaturant- and force-induced unfolding pathways are clearly different, despite having similar barriers to unfolding. A full Φ-value analysis of the protein in bulk solution leads to the same conclusion: partial Φ-values for residues in the A strand ($\Phi_{I2A} = 0.45$, $\Phi_{V4A} = 0.29$) indicate that it is still structured in the transition state, which is inconsistent with the simulated mechanism of forced unfolding.[57]

The mechanical unfolding of barnase, mentioned above, is an experimental test for the similarity of unfolding pathways, having already been thoroughly studied using chemical denaturation. As described above, barnase was inserted into modules 2, 4 and 6 of the titin I27 construct. To check that the construct had no effect on bulk stability or unfolding kinetics, the unfolding rate and stability of both I27 and barnase were checked in a three-module (I27-barnase-I27) construct, and found to have the same values as the isolated proteins, within error. ^{1}H-^{15}N HSQC NMR spectra of the modules in isolated form and in the construct could be overlaid almost perfectly, demonstrating that no structural alteration had taken place either. Fig. 6 (a) shows the dependence of unfolding force on

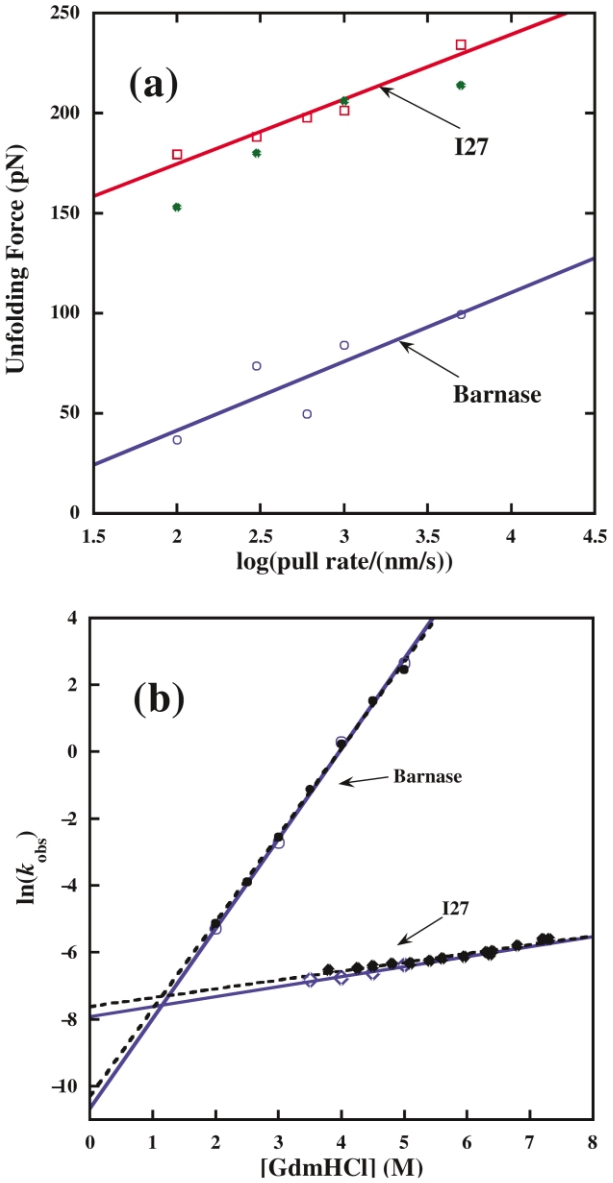

Fig. 6 Unfolding of barnase and titin I27 by AFM and chemical denaturant. (a) Dependence of unfolding force on pulling speed and (b) extrapolation of unfolding rate in bulk solution to zero denaturant. (Taken from reference 28.)

pulling speed for barnase and I27, from which unfolding rates at zero force are calculated from Monte-Carlo simulations to be 6.7×10^{-2} s^{-1} and 3.0×10^{-4} s^{-1} for barnase and I27 respectively, while Fig. 6 (b) shows the unfolding rates of each protein extrapolated to 0 M denaturant. While barnase unfolds an order of magnitude more slowly than I27 at zero denaturant, it unfolds about two orders of magnitude faster at zero force—so barnase must unfold by different pathways under the different conditions.[28] Again, this is borne out by molecular dynamics simulations, which show that the transition state for mechanical unfolding of barnase is much more native-like than that for thermal unfolding, with more secondary and tertiary structure being preserved. As for titin I27, the effect of force is

to unravel the protein from the termini, while temperature or denaturant cause a more global expansion and rendering of the structure (Fig. 7).

It must be pointed out that, because of computational limitations, molecular dynamics simulations of mechanical unfolding are in a very different regime from the experiments, with pulling speeds around six orders of magnitude faster. However, within the range of pulling speeds that have been accessed by MD simulation (several orders of magnitude of pulling speed), the unfolding pathway is consistent, lending confidence that the results will still hold at the much slower experimental speeds.[28] The difficulty of time scale is one that is also faced in comparing molecular dynamics simulation with bulk unfolding data in denaturant.[2] Nonetheless, it is the only method able to directly furnish atomistic detail on the unfolding process, and the results have been shown to be in both qualitative and semi-quantitative agreement with experiments.[13]

Probing unfolding pathways in detail by AFM experiments

The explosive unfolding of most proteins in AFM experiments limits the amount of detail that can be experimentally inferred about unfolding pathways. Nonetheless, to use the new technique to its full potential for investigating unfolding mechanisms, one would like to be able to localize the interactions that are critical in the unfolding pathway. Some experiments have already been interpreted at a more microscopic level. Force-mode AFM on a tetramer of the α-spectrin R16 domain was used to infer an intermediate, based on the distribution of contour length increases:[19] a bimodal distribution of extension was obtained, in which the shorter length was attributed to the intermediate (the interpretation is not entirely clear as the shorter extensions only occurred for the first peaks in any trace, despite all the modules being identical).

The intermediate seen in the molecular dynamics simulations of I27 has been invoked to explain the shoulder on the back of each unfolding peak in a poly-I27 construct.[58] This is rationalized as an increase in contour length upon detachment of the A strand. Force-extension traces for the A strand mutant K6P showed no shoulder, presumably because the A strand was initially detached or detached at very low force. Further experiments making mutations in the A strand allowed the change in unfolding rate constant on mutation to be evaluated from Monte-Carlo simulation. However, as the mutants were all to proline, the results are difficult to interpret: as intended, proline cannot act as a backbone hydrogen bond donor, but it also imposes tight geometric restrictions on backbone conformation that may disrupt the structure and side chain packing.

To check whether the structure of the proposed titin unfolding intermediate is reasonable, a mutant was expressed with the A strand deleted: the loss of stability, 2.78 kcal.mol^{-1}, was only slightly greater than that caused by single mutations to the A strand, and the overall stability was still high (4.82 kcal mol^{-1}). Structural integrity was confirmed by ^{1}H-^{15}N HSQC NMR spectra of the mutant, which showed only localized chemical shift changes. A non-disruptive mutation to the A strand caused no change in mechanical unfolding properties. The most plausible interpretation of this is that unfolding occurs from an intermediate in which the A strand is detached.[59]

To take this further, we can draw on methods already developed for analyzing folding kinetics in bulk experiments with chemical denaturant. A good method for quantitatively mapping the participation of each residue in the transition-state is Φ-value analysis.[43,44] A Φ-value analysis for folding measures the change in activation free energy for folding, $\Delta\Delta G_{D-TS}$, as a fraction of the change in stability, $\Delta\Delta G_{D-N}$, after a single point mutation ($\Phi_F = \Delta\Delta G_{D-TS}/\Delta\Delta G_{D-N}$). Then a mutated residue making all its native contacts (*i.e.* in a fully formed region of the structure) in the transition state would have a Φ-value for folding, Φ_F, of 1.0, since the mutation destabilizes

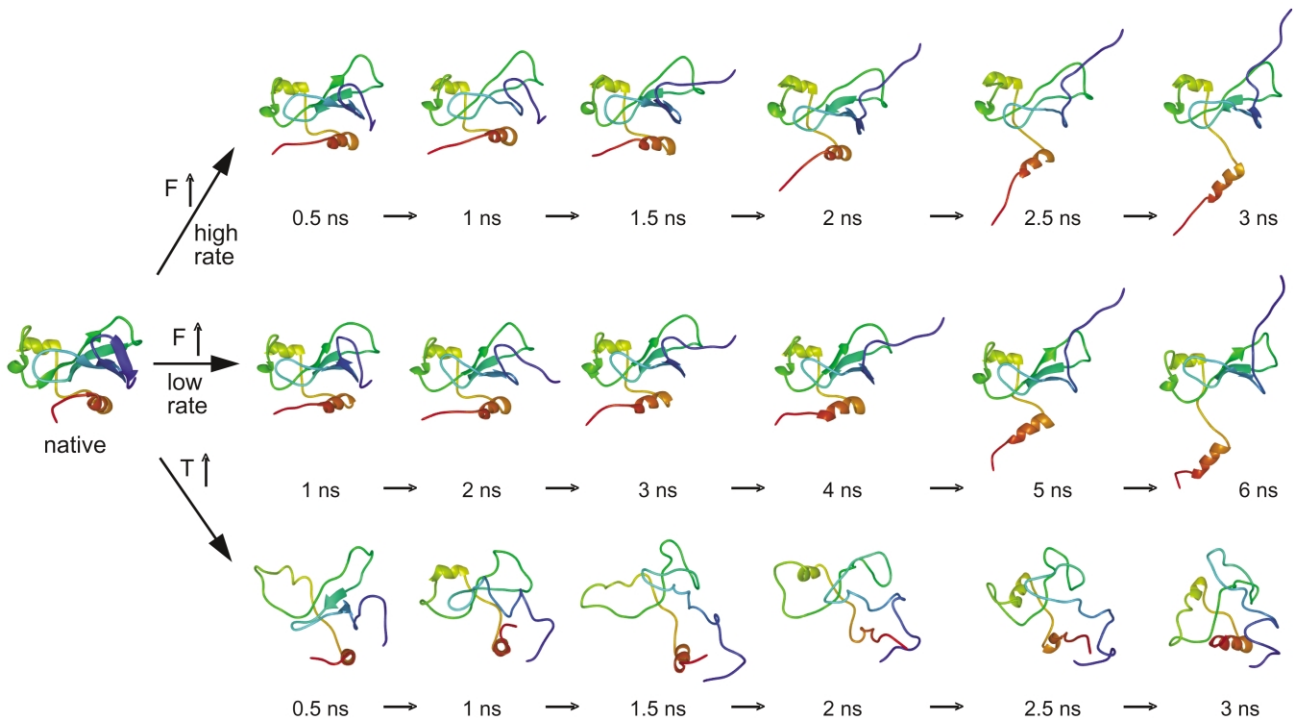

Fig. 7 Unfolding pathway of barnase in MD simulations by force and by temperature. Three unfolding pathways are illustrated: a mechanical unfolding pathway at 'high' pulling rate (pulling speed of 0.02 Å/ps, *top*), one at 'low' pulling rate (pulling speed of 0.01 Å/ps, *middle*), and a thermal unfolding pathway at 498 K (*bottom*). The two mechanical pathways have similar features (although different time-scales), showing that the mechanical unfolding pathway does not depend strongly on pulling speed. The thermal unfolding pathway differs substantially from these, taking place *via* a global disruption of the structure, compared with the sequential unraveling in the mechanical unfolding. (Taken from reference 28.)

both transition state and native state equally, while a residue in an unstructured part of the protein would have a Φ_F of 0. If the folding is two-state, this value can be obtained from the unfolding activation free energy, since $\Delta\Delta G_{D-TS} = \Delta\Delta G_{D-N} - \Delta\Delta G_{TS-N}$. The change in stability can be directly obtained as the difference of wild-type and mutant energies. The change in unfolding activation free energy is given by $\Delta\Delta G_{TS-N} = -RT\ln(k_u^{wt}/k_u^{mut})$ from transition state theory.[48]

We propose that a protein engineering method similar to Φ-value analysis could be used to analyze mechanical protein (un)folding pathways by AFM. The most obvious way of doing this would be to use unfolding rate constants as obtained from the match of Monte-Carlo simulations to experimental data to calculate $\Delta\Delta G_{TS-N}$, the justification for this being the same as for conventional Φ-values. Since a change in stability under force cannot be measured by AFM (it would require pulling many orders of magnitude slower than is possible), one could take the change in stability from conventional equilibrium denaturation experiments. Some caution is needed here, though: firstly, one needs to be sure that the unfolding observed in the AFM experiment occurs from the native state, which is not the case for titin I27. Secondly, the 'denatured state' measured by AFM is clearly different from that in solution, being highly extended. However, it should still be possible to use stability changes measured by conventional means, if the changes in stability upon mutation are confined to the native state, as is usually assumed. The nature of mutations is also important, just as in bulk experiments. The effect of an arbitrary mutation is very difficult to predict or interpret, since there are many contributions to the free energy: solvation, electrostatic effects, dispersion forces and chain entropy can all play a part. It is most useful to make conservative mutations, whose effect can be simply interpreted, for example side-chain deletion mutants in the hydrophobic core which preserve charge. In this case, the dominant contribution to free energy changes is from hydrophobic packing interactions, in which case the Φ-values can be understood, to a good approximation, as a fraction of native contacts in the transition state.

The situation may be complicated if there are mechanical unfolding intermediates, as in titin I27. There is no way to measure rigorously the stability change from the intermediate to denatured state by experiment, and a picture of the unfolding pathway can then only be assembled by using experimental results and atomistic simulations in conjunction with each other.[59]

Conclusions

Much has already been learnt about protein folding from theory, simulation and experiment, leading to a 'new view' of folding as taking place on a multidimensional energy landscape, funneled toward the native state. Experiments using chemical denaturants or temperature are one method of probing the energy landscape, which may be physiologically appropriate for many globular proteins (such as enzymes). Force allows us to unfold proteins in a fundamentally different way, which should be more suitable for proteins with mechanical function, such as muscle proteins or cell adhesion molecules. Even for proteins that do not have a mechanical function, force provides another method of studying folding to complement more traditional approaches. The most basic parameters resulting from an AFM experiment already describe the unfolding potential to some extent: a distance to the transition state, x_u, which depends on the elasticity of the folded protein, and an unfolding rate at zero force, k_u^o, that can be compared with data from bulk solution studies. Mechanically robust proteins will have small x_u and k_u^o.

AFM data on the mechanical unfolding of proteins from a wide variety of sources are now available. These show that some classes of proteins, such as immunoglobulin-like β-sandwich domains, are much more mechanically robust than others, such as the all-helical spectrin. Hydrogen bonding between β-strands, which holds together the β-sandwich domains is not present in the helical proteins and is probably the primary reason for the difference in unfolding forces for these classes of protein. Within the β-sandwich fold, it appears that the topology of the immunoglobulin superfamily is more robust than that of the fibronectin type III superfamily. However, both folds are able to accommodate a wide range of mechanical stability to adopt different functions, so the sequence and tertiary interactions can modulate this strength. Non-mechanical proteins unfold at low forces, as might be expected, and like spectrin, force is mainly resisted by tertiary interactions.

A combination of molecular dynamics simulations and experiments using AFM and conventional folding techniques has been used to look at unfolding pathways under force, at high temperature and in denaturant. It appears from the simulations that the force-induced and thermal unfolding pathways are always very different. The transition state under force is much more native-like than that at high temperature. Unfolding by force occurs by sequential unraveling of the structure from the termini, while temperature causes a global expansion and disruption of structure. Experimental results for the mechanical unfolding of barnase also show very different unfolding rates at zero force and zero denaturant, supporting the idea of different pathways. Intriguingly, though, unfolding rate constants for titin I27[26a] determined by the different methods are within error of each other and the same is true of I28;[27] perhaps this property is particular to mechanical proteins.

The future of studying protein folding by AFM lies in going beyond the study of different wild-type proteins, to investigate the mechanical unfolding pathways in detail. In addition to the methods already used, a protein engineering approach will be very informative. Using this, the key interactions important for mechanical stability can be identified and related to the predictions made by simulations. It is expected that the relative importance of tertiary interactions will become more apparent when these methods are applied to structurally related proteins. Finally, it will be possible to compare the folding pathways and folding nucleus determined in this way with that from conventional Φ-value analysis.

Acknowledgements

We thank Emanuele Paci, Martin Karplus, José Luis Toca-Herrera, Kathryn Scott and Susan Fowler for helpful discussions, and Valerie Daggett and Bin Li for providing Fig. 7 and the cover picture. These two figures and Fig. 6 are reproduced here with the kind permission of the Biophysical Society. Robert Best is supported by a Mandela Scholarship from the Cambridge Commonwealth Trust and Jane Clarke is a Wellcome Trust Senior Research Fellow. This work was funded by the Wellcome Trust and the Medical Research Council.

Notes and references

† Note that the values of ν and κ are likely to be very different from the those for chemical reactions, but the argument here does not depend on knowing them.[48]

1 C. B. Anfinsen, *Science*, 1973, **181**, 223.
2 U. Mayor, C. M. Johnson, V. Daggett and A. R. Fersht, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, 13518.
3 J. Balbach, V. Forge, N. A. J. v. Nuland, S. L. Winder, P. J. Hore and C. M. Dobson, *Nat. Struct. Biol.*, 1995, **2**, 865.
4 C. Levinthal, *J. Chim. Phys.*, 1968, **65**, 44.
5 J. D. Bryngelson, J. Onuchic, N. D. Socci and P. G. Wolynes, *Proteins: Struct., Funct., Genet.*, 1995, **21**, 167.
6 J. N. Onuchic, H. Nymeyer, A. E. Garcia, J. Chahine and N. D. Socci, *Adv. Prot. Chem.*, 2000, **53**, 87.
7 D. J. Wales, M. A. Miller and T. R. Walsh, *Nature (London)*, 1998, **394**, 758.
8 J. D. Bryngelson and P. G. Wolynes, *Proc. Natl. Acad. Sci. U.S.A.*, 1987, **84**, 7524.

9 J. D. Bryngelson and P. G. Wolynes, *J. Phys. Chem.*, 1989, **93**, 6902.
10 L. Mirny and E. Shakhnovich, *Annu. Rev. Biophys. Biomol. Struct.*, 2001, **30**, 361.
11 D. K. Klimov and D. Thirumalai, *J. Chem. Phys.*, 1998, **109**, 4119.
12 J.-E. Shea, J. N. Onuchic and C. L. B. III, *J. Chem. Phys.*, 2000, **113**, 7663.
13 V. Daggett, A. Li and A. R. Fersht, *J. Am. Chem. Soc.*, 1998, **120**, 12740.
14 T. Lazaridis and M. Karplus, *Science*, 1997, **278**, 1928.
15 G. Binnig, C. F. Quate and C. Gerber, *Phys. Rev. Lett.*, 1986, **56**, 930.
16 E. Evans, K. Ritchie and R. Merkel, *Biophys. J.*, 1995, **68**, 2580.
17 M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez and H. E. Gaub, *Science*, 1997, **276**, 1109.
18 M. Rief, M. Gautel, A. Schemmel and H. E. Gaub, *Biophys. J.*, 1998, **75**, 3008.
19 P.-F. Lenne, A. J. Raae, S. M. Saraste and J. K. H. Horber, *FEBS Lett.*, 2000, **476**, 124.
20 M. Rief, J. Pascual, M. Saraste and H. E. Gaub, *J. Mol. Biol.*, 1999, **286**, 553.
21 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. U.S.A.*, 2000, **97**, 139.
22 A. A. Deniz, T. A. Laurence, G. S. Beligere, M. Dahan, A. B. Martin, D. S. Chemla, P. E. Dawson, P. G. Schultz and S. Weiss, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, 5179.
23 X. Zhuang, T. Ha, H. D. Kim, T. Centner, S. Labeit and S. Chu, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, 14241.
24 A. S. Dakkouri, D. M. Kolb, R. Edelstein-Shima and D. Mandler, *Langmuir*, 1996, **12**, 2849.
25 J. L. Hutter and J. Bechhoefer, *Rev. Sci. Instrum.*, 1993, **64**, 1868.
26 (*a*) 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. U.S.A.*, 1999, **96**, 3694; (*b*) A. F. Oberhauser, P. E. Marszalek, M. Carrion-Vazquez and J. M. Fernandez, *Nat. Struct. Biol.*, 1999, **6**, 1025.
27 H. Li, A. F. Oberhauser, S. B. Fowler, J. Clarke and J. M. Fernandez, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, 6527.
28 R. B. Best, B. Li, A. Steward, V. Daggett and J. Clarke, *Biophys. J.*, 2001, **81**, 2344.
29 E. Evans, *Annu. Rev. Biophys. Biomol. Struct.*, 2001, **30**, 105.
30 M. B. Viani, T. E. Schäffer, A. Chand, M. Rief, H. E. Gaub and P. K. Hansma, *J. Appl. Phys.*, 1999, **86**, 2258.
31 M. B. Viani, T. E. Schäffer, G. T. Paloczi, L. I. Pietrasanta, B. L. Smith, J. B. Thompson, M. Richter, M. Rief, H. E. Gaub, K. W. Plaxco, A. N. Cleland, H. G. Hansma and P. K. Hansma, *Rev. Sci. Instrum.*, 1999, **70**, 4300.
32 E. Evans and K. Ritchie, *Biophys. J.*, 1997, **72**, 1541.
33 E. Evans and K. Ritchie, *Biophys. J.*, 1999, **76**, 2439.
34 A. F. Oberhauser, P. E. Marszalek, H. P. Erickson and J. M. Fernandez, *Nature (London)*, 1998, **393**, 181.
35 Y. Oberdorfer, H. Fuchs and A. Janshoff, *Langmuir*, 2000, **16**, 9955.
36 P. Carl, C. H. Kwok, G. Manderson, D. Speicher and D. E. Discher, *Proc. Natl. Acad. Sci. U.S.A.*, 2001, **100**, 1.
37 A. Bateman, E. Birney, R. Durbin, S. R. Eddy, R. D. Finn and E. L. L. Sonnhammer, *Nucleic Acids Res.*, 1999, **27**, 260.
38 K. A. Scott, A. Steward, S. B. Fowler and J. Clarke, *J. Mol. Biol., in press*.
39 C. M. Goll, A. Pastore and M. Nilges, *Structure (London)*, 1998, **6**, 1291.
40 P. Amodeo, F. Fraternali, A. M. Lesk and A. Pastore, *J. Mol. Biol.*, 2001, **311**, 283.
41 J. G. Head, A. Houmeida, P. J. Knight, A. R. Clarke, J. Trinick and R. L. Brady, *Biophys. J.*, 2001, **81**, 1570.
42 N. C. Gassner, W. A. Baase, J. D. Lindstrom, J. Lu, F. W. Dahlquist and B. W. Matthews, *Biochemistry*, 1999, **38**, 14451.
43 L. Serrano, A. Matouschek and A. R. Fersht, *J. Mol. Biol.*, 1992, **224**.
44 A. Matouschek, L. Serrano and A. R. Fersht, *J. Mol. Biol.*, 1992, **224**, 819.
45 N. D. Socci, J. N. Onuchic and P. G. Wolynes, *Proc. Natl. Acad. Sci. U.S.A.*, 1999, **96**, 2031.
46 D. K. Klimov and D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.*, 1999, **96**, 6166.
47 M. Rief, J. M. Fernandez and H. E. Gaub, *Phys. Rev. Lett.*, 1998, **81**, 4764.
48 A. R. Fersht, *Structure and mechanism in protein science*, 1st edn. 1999, W. H. Freeman, New York.
49 G. I. Bell, *Science*, 1978, **200**, 618.
50 A. F. Oberhauser, P. K. Hansma, M. Carrion-Vazquez and J. M. Fernandez, *Proc. Natl. Acad Sci. U.S.A.*, 2001, **98**, 468.
51 E. Paci and M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, 6521.
52 E. Evans, *Faraday Discuss.*, 1998, **111**, 1.
53 E. Evans, A. Leung, D. Hammer and S. Simon, *Proc. Natl. Acad. Sci. (U.S.A.)*, 2001, **98**, 1.
54 H. Lu, B. Isralewitz, A. Krammer, V. Vogel and K. Schulten, *Biophys. J.*, 1998, **75**, 662.
55 H. Lu and K. Schulten, *Chem. Phys.*, 1999, **247**, 141.
56 H. Lu and K. Schulten, *Biophys. J.*, 2000, **79**, 51.
57 S. B. Fowler and J. Clarke, *Structure*, 2001, **9**, 355.
58 P. E. Marszalek, H. Lu, H. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten and J. M. Fernandez, *Nature (London)*, 1999, **402**, 100.
59 S. B. Fowler, R. B. Best, T. J. Rutherford, A. Steward, E. Paci, M. Karplus and J. Clarke, *unpublished data*.