Single-Molecule Studies of Protein Folding

Alessandro Borgia,[1] Philip M. Williams,[2] and Jane Clarke[1]

[1]Department of Chemistry, Cambridge University, Medical Research Council Centre for Protein Engineering, Cambridge, CB2 1EW, United Kingdom; email: ab697@cam.ac.uk, jc162@cam.ac.uk

[2]Laboratory of Biophysics and Surface Analysis, School of Pharmacy, University of Nottingham, Nottingham NG7 2RD, United Kingdom; email: phil.williams@nottingham.ac.uk

Annu. Rev. Biochem. 2008. 77:101–25

First published online as a Review in Advance on April 15, 2008

The *Annual Review of Biochemistry* is online at biochem.annualreviews.org

This article's doi: 10.1146/annurev.biochem.77.060706.093102

0066-4154/08/0707-0101$20.00

Key Words

AFM, energy landscape, folding kinetics, FRET, protein dynamics

Abstract

Although protein-folding studies began several decades ago, it is only recently that the tools to analyze protein folding at the single-molecule level have been developed. Advances in single-molecule fluorescence and force spectroscopy techniques allow investigation of the folding and dynamics of single protein molecules, both at equilibrium and as they fold and unfold. The experiments are far from simple, however, both in execution and in interpretation of the results. In this review, we discuss some of the highlights of the work so far and concentrate on cases where comparisons with the classical experiments can be made. We conclude that, although there have been relatively few startling insights from single-molecule studies, the rapid progress that has been made suggests that these experiments have significant potential to advance our understanding of protein folding. In particular, new techniques offer the possibility to explore regions of the energy landscape that are inaccessible to classical ensemble measurements and, perhaps, to observe rare events undetectable by other means.

Contents

INTRODUCTION

The "protein-folding problem" comprises three fundamental questions. The first is essentially a question of thermodynamics: How does the amino acid sequence specify the three-dimensional (3D) structure? The second is a question of kinetics: How does the polypeptide chain attain its native state (N) in a biologically relevant timescale? Finally, how do proteins avoid misfolding? Despite recent advances, it is still hard to predict structure from sequence, particularly where there are no homologues; progress toward predicting folding pathways is significant, but many features of folding mechanisms remain unclear. Moreover, designing proteins de novo with specific structures and/or specific properties, is still nontrivial, although there have been some notable successes (1).

One of the most interesting features of evolved proteins is their cooperative nature. At the gross level, they fold and unfold cooperatively in an all-or-none fashion. At equilibrium, the vast majority of small, single domain proteins fold in a two-state manner; that is, all molecules exist in either folded or denatured

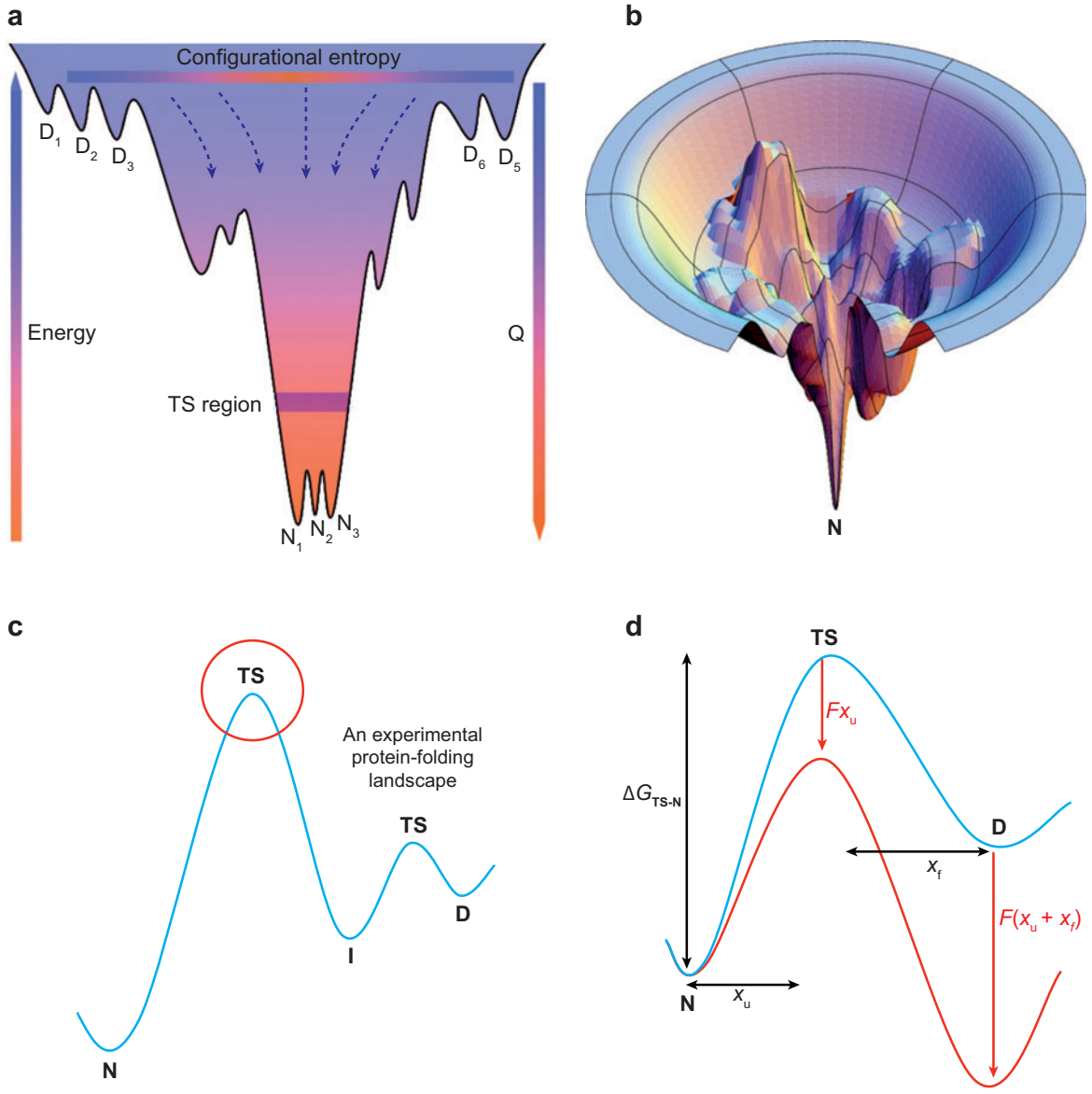
N: native state

Figure 1

Funnel-shaped energy landscapes for proteins in folding conditions. (*a*) Cross section of a protein-folding energy landscape. The case represented refers to a protein with considerable ruggedness in the denatured state (D) that folds to a native state (N), which has three slightly different conformations. A true protein-folding funnel is a multidimensional object in which a large number of parameters need to be taken into account for an exhaustive description of the protein at every point of the landscape. For the sake of simplicity, the system is described here by the configurational entropy term on the x-axis and two dimensions for the y-axis: the energy of the conformation (i.e., the free energy with the chain entropy term subtracted) and the fraction of native contacts, or "degree of nativeness," Q, as in Reference 129. (*b*) Three-dimensional (3D) pictorial representation of a rugged energy landscape: a wide ensemble of denatured states (D) populating the external rim of the energy landscape (great conformational entropy, high energy) can flow through many alternative routes toward the native state (N) at the bottom (lowest energy, lowest conformational entropy). The fraction of molecules folding through each pathway are the result of a statistical selection, depending on the relative heights of the energy barriers en route to the native state on that pathway. (Picture reproduced with permission from Reference 2.) (*c*) An "experimental" energy landscape. In classical protein-folding experiments, the relative stabilities of the N, D, and populated intermediate (I) states can be determined, as well as folding and unfolding rate constants, which depend upon the differences in free energy between the starting state and the transition state (TS). (*d*) How force affects the energy landscape. The TS and D state are lowered in energy by force. The change in energy on applying a force, F, is the product of F and the distance from the native state (N) (x_u and $x_u + x_f$).

form. Partly structured intermediates are unstable and do not have long lifetimes. An important concept underlying our understanding of protein folding is that for proteins to fold rapidly, they must have funnel-shaped energy landscapes (**Figure 1**) (2, 3). Evolved proteins are understood to have relatively smooth energy landscapes, essentially lacking kinetic traps (stable partly folded intermediates). Many experimental studies are devoted to characterizing structures on this energy landscape. The nature of a protein-folding energy landscape suggests that although the "flux" of proteins may have average features, any single molecule may take any of a number of routes through that landscape; thus, the concept of there being multiple folding pathways has become important. What is the nature of kinetic barriers to folding? Do some proteins fold downhill—without any barriers

Energy landscape: the energy of the polypeptide chain for each conformation it can assume between D and N, as a function of the conformational freedom



DSE: denatured state ensemble

smFRET: single-molecule Förster resonance energy transfer

to folding? What are the features of molecules within any rapidly interchanging ensemble, such as the denatured state ensemble (DSE)? What is the structure of intermediates along the folding pathways? Are they obligatory "on-pathway" species or are they "off pathway"? All these questions are of current importance and interest.

Traditional experimental approaches have been extraordinarily powerful at answering many of these questions. In particular, protein engineering phi-value analysis (4), advances in NMR spectroscopy (5), and computational methods (6) have been vital. But traditional biophysical experiments can only tackle ensemble measurements—we can only observe average properties of $\sim 10^{14}$–10^{17} molecules at a time. In principle, single-molecule protein-folding experiments should allow us to resolve the properties of individual molecules and quantify subpopulations within, say, the denatured ensemble or to observe molecules folding by alternative pathways, for instance. In this review, we discuss progress toward this goal.

USING SINGLE-MOLECULE FLUORESCENCE TO STUDY PROTEIN FOLDING

Spectroscopic techniques, in particular circular dichroism, fluorescence, and NMR have provided crucial insights into the mechanism by which a polypeptide chain attains its native fold (7, 8). Fluorescence undoubtedly plays a pivotal role. It is the process by which an electron, promoted to a higher energy level on interaction with a photon of given energy (in the 10^{-15} s time regime), returns to its ground state $\sim 10^{-8}$ s later, dissipating the acquired energy via emission of a photon of lower energy. This relaxation phenomenon is extremely dependent on environmental factors, and the properties of the resulting emitted fluorescence (wavelength, quantum efficiency, lifetime) are exquisite reporters of changes in the environment of the fluorophore (9). Moreover, because the excited state is relatively long lived compared to the time frame of the electronic excitation (nanoseconds, ns, versus femtoseconds, fs), other radiation-less phenomena compete with the relaxation by fluorescence, increasing the number of ways in which this spectroscopic tool may be exploited to gain structural information. FRET (Förster resonance energy transfer) is one of these phenomena and is particularly useful to study protein folding. FRET between two fluorophores occurs when an excited electron of one fluorophore, the donor (D), transfers its energy to a second fluorophore, the acceptor (A), providing the emission spectrum of D and the absorption spectrum of A overlap to a suitable extent and the distance between them is within a given range (10).

The reasons for the success of FRET at the single-molecule level are its inherent sensitivity and the steep dependence of the efficiency of the energy transfer (E_{ET}) on the interfluorophore distance d_{DA}, [$E_{ET} \propto (d_{DA})^{-6}$]. Even subtle conformational changes in the protein translate into a measurable change in E_{ET}, and the distance-resolution of the technique (10–100 Å) is appropriate for monitoring conformational changes in proteins. A complication arises from the fact that E_{ET} is not only a function of the interdye distance, but it also depends on the angles between the respective dipoles and between their associated transition moments. Fortunately though, in most cases, this does not compromise its applicability to protein-folding studies (11). Another considerable disadvantage is that, despite the natural occurrence of tryptophan in proteins, this cannot be used in single-molecule FRET experiments (smFRET); even if endowed with a reasonable quantum yield, the problem is its intrinsic photolability. Thus, the use of extrinsic organic dyes is compulsory, and the commercial availability of an increasing number of fluorophores is benefiting the applicability of smFRET studies to the field. However, the labeling procedure is still a barrier to the general use of FRET techniques, mainly because it may be very time consuming (e.g., identification of adequate attachment sites,

protein engineering, dye attachment). Moreover, the large size of the dyes and the potential effects that this may have on the stability and the folding of the protein makes thermodynamic and kinetic characterization of all the labeled species strictly necessary, even for the most carefully designed labeled proteins. There have been a number of excellent recent reviews describing FRET technology, so we will not discuss the methodology or instrumentation in more detail here (12–15).

Equilibrium Measurements

As in traditional protein-folding studies, smFRET experimens have been used to investigate both equilibrium and kinetic properties of proteins.

Observing equilibrium populations. In the simplest case of a smFRET equilibrium experiment, the laser of a confocal apparatus is focused on a dilute protein solution (10–100 pM), to obtain a focal volume of few femtoliters. This maximizes the probability of observing only one "in focus" molecule at a time. When a freely diffusing double-labeled molecule traverses the confocal volume, the laser beam will excite (almost exclusively) the donor dye. If the above-mentioned conditions for FRET between the two fluorophores are fulfilled, the donor will transfer its excitation energy to the acceptor via the Förster mechanism. Bursts of photons of single excited molecules are recorded and integrated for the few milliseconds (ms) the protein spends in the confocal volume; this is repeated thousands of times in order to obtain meaningful statistics. The intensity of both acceptor and donor fluorescence emission is measured, and used to build histograms of E_{ET} distributions, which reflect the interdye distances. The number of peaks in such distributions represents subpopulations of different protein conformations resolved in the given conditions, whereas the corresponding mean values indicate the average interdye distances within those populations [and thus the compactness of the polypeptide chain (16, 17)]. Another potentially useful characteristic of each state is given by the width of the distribution, which is related to the low degrees of freedom of the protein in that state. However, a quantitative analysis of the width is complicated by the existence of other incompletely understood sources of broadening and has remained controversial (18–20). Studies of two-state proteins at equilibrium (18–21) have highlighted a bimodal distribution of the E_{ET}, whose relative mean values correspond to either the fully folded or the fully unfolded populations only (**Figure 2**). In favorable cases, it is possible to determine the free energy of the protein from the populations of molecules in the two ensembles (18). Particularly interesting is a recent study of the B domain of protein A (21) in which the bimodal distribution in conjunction with the absence of any shift in the distribution of mean value of E_{ET} of the N ensemble upon addition of denaturant was used to support a cooperative folding mechanism, rather than a "downhill folding" scenario, which might be predicted for such superfast folding proteins (22).

It is interesting to note that these studies demonstrate that the denatured state observed under folding conditions, at low denaturant concentrations, is significantly more collapsed than at higher denaturant concentrations (**Figure 2**) (23, 24). Schuler and coworkers (19), who have investigated the folding of a small fast folding cold-shock protein, Csp*Tm*, demonstrated that the denatured state compactness shows a continuous, noncooperative decrease with increasing guanidinium chloride (GdmCl) concentration. Eaton and coworkers (20) have investigated this process further, successfully combining smFRET measurements with molecular dynamics simulations, demonstrating the power of this approach for elucidating structural properties of unfolded proteins in detail. In a recent study, using multiple dye pairs in different parts of the protein, Schuler and coworkers have described the solvent-dependent chain contraction of Csp*Tm* as an isotropic transition,

Förster mechanism: the singlet-singlet electronic energy transfer (not fluorescence transfer) between two molecules, based on long-range dipole-dipole resonance coupling

Downhill folding: reaction in which the denatured state reaches the native conformation through a very small or nonexistent energy barrier

Collapse: very fast contraction of the polypeptide chain occurring in response to a sudden change in solvent conditions

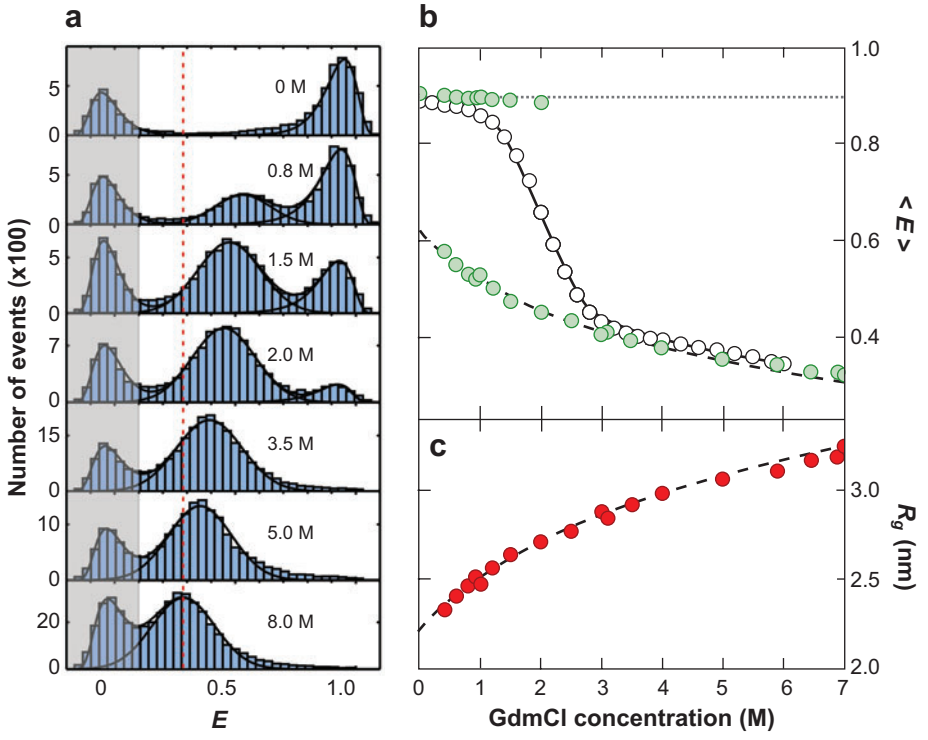


Figure 2

FRET studies of protein folding at equilibrium. (*a*) Transfer efficiency histograms of Csp*Tm* labeled with Alexa 488 and 594 at the chain termini at several GdmCl concentrations. The peak at high *E* corresponds to folded protein, the peak at lower *E* to unfolded protein. The signal at $E \approx 0$ originates from molecules with inactive acceptor. (*b*) Mean transfer efficiencies $\langle E \rangle$ of folded (*top*, *blue circles*, corrected for refractive index change) and unfolded (*bottom*, *blue circles*) subpopulations from single-molecule measurements compared to an ensemble unfolding transition of the same sample (*pink circles*) with a two-state fit (*purple*) (*c*) Radius of gyration (R_g) calculated from the $\langle E \rangle$ of the unfolded chain assuming the distance distribution of a Gaussian chain, showing the collapse of the unfolded state. Dashed lines are empirical fits to the equivalent of a binding model. This figure was kindly provided by Ben Schuler. Data taken from References 19, 25, and 44.

suggesting that the collapse of this protein is a nonspecific process, evenly distributed across the polypeptide chain (25), rather than driven by the early formation of specific clusters of interactions, as has been suggested for other proteins (26–30).

Observing polypeptide dynamics at equilibrium. A number of classical ensemble experiments, in particular NMR spectroscopy, have demonstrated that the denatured states of some proteins have a significant residual structure (see References 5, 31, and 32 for recent reviews). Whether this residual structure plays a role in determining protein-folding pathways is perhaps controversial. If interconversion between species in the DSE is much faster than the overall folding process, then it seems unlikely that DSE heterogeneity is the origin of parallel protein-folding pathways (33–37). SMFRET offers the opportunity to investigate the dynamics of the denatured state.

Eaton and coworkers used single-molecule FRET to characterize dynamics in the DSE of denatured Csp*Tm* (19). The E_{ET} distribution of a double-labeled Csp*Tm* was recorded

under unfolding conditions and compared with the distribution obtained in the same conditions for rod-like double-labeled polyproline chains of different lengths, models of a rigid, fully extended state (38). The results suggested that the reconfiguration time was much more rapid than the observation time. Thus, it was possible to place an upper limit on the reconfiguration time of the denatured polypeptide chain of ~25 microseconds (μs). The authors were able to use this limit to estimate the elusive preexponential factor ($2\pi\tau_0$) that represents the maximum folding time in a barrier-free scenario. According to a simplified version of the Kramers' theory, the folding time (τ_F) of a single molecule is given by $2\pi\tau_{0*}e^{(\Delta G^{TS}/k_BT)}$ (39–41). For Csp*Tm*, τ_0 was estimated to be between 40 ns and 25 μs; this value, together with the measured folding time of 12 ms, was used to determine lower and upper limits for activation energy barriers for folding ($11\,k_BT < \Delta G^{TS} < 4\,k_BT$). These results represent the first experiment-based calculation of the range of values of an energy barrier separating two states lying in consecutive energy wells, providing a new benchmark for theoretical calculations of free-energy surfaces.

Similarly, Nienhaus and coworkers (42) studied immobilized RNaseH and estimated a τ_0 of 20 μs, yielding the surprising value of 10 k_BT, (or 170 ns and 15 k_BT, according to the diffusion limits of a Gaussian chain) for the height of the barriers between the interconverting conformations in the DSE in 6 *M* GdmCl. Such high barriers suggest an unexpected roughness of the energy landscape, even in the presence of high denaturant concentrations.

Even after significant instrumental improvements, the investigation of phenomena occurring in the submicrosecond timescale with FRET remains problematic. With the discovery of ultrafast-folding proteins (k_f = 10^5–10^6 s^{-1}) (43), it became increasingly important to shed light on the features of the polypeptide chain dynamics of the denatured state in folding conditions. Schuler and coworkers (44) addressed this issue by analyzing the photon statistics of an N- and C-termini labeled Csp*Tm*. Employing subpopulation-specific correlation methods and recording the time intervals between consecutive detected photons with picosecond (ps) resolution, they could determine the relative intramolecular diffusion coefficients for the donor and acceptor at the ends of the unfolded chain. The data were fitted to a diffusive chain dynamics model and the corresponding reconfiguration time of the protein was obtained in various denaturant concentrations. Upon protein collapse, the authors highlighted a progressive deceleration of the chain dynamics, dependent on the denaturant concentration, which they suggest could indicate an increase of the "internal friction" in the collapsed DSE; whereas the collapse itself is kinetically consistent with a barrierless, purely diffusive process, contrary to what has been described in other cases (23, 26, 45, 46). With the ability of this technique to monitor single-molecule dynamics in the collapsed or denatured state even under folding conditions and with ns time resolution, all the power of smFRET, as yet unexploited, is captured.

In the energy landscape scenario, a greater ruggedness implies slower interconversion between the DSE conformations, thus creating the preconditions for elements of structure to form in a persistent fashion (so-called residual structure). In this respect, it is perhaps not a coincidence that Nettels et al. detected no energy barriers for Csp*Tm* in unfolding conditions (44) and described the following solvent-dependent chain contraction as a downhill, nonspecific process; whereas Nienhaus et al. (42), who estimated a value of ~12.5 k_BT for RNaseH, proposed a barrier-limited collapse, which is plausibly biased by preorganized clusters of interactions. It will be interesting to verify if the proteins, for which residual structure in D has been detected or inferred, actually sample a rugged energy surface.

Preexponential factor: the coefficient of proportionality between the reaction rate constant and the energy barrier, $k_r = A\exp(-\Delta G^{TS}/k_BT)$. Consequently, when $\Delta G^{TS} = 0 \rightarrow k_r = A$.

Kinetic Measurements

In liquid environments, the polypeptide chain moves as a result of thermal agitation (Brownian motion) at a speed given by its typical diffusion coefficient, with the result that the time a molecule spends in the detection volume of a typical free diffusion smFRET experiment is a few ms. This may represent a minor problem in equilibrium experiments, but it is obviously a severe limitation in kinetic experiments because many proteins fold over a time of seconds to minutes in physiological conditions. Several strategies were studied to overcome this problem (47–51), but few have been demonstrated as successful.

Extending the observation time. At a first glance, the most straightforward way to extend the observation time would involve the immobilization of the protein onto a suitable surface. Among the variety of solutions proposed, that developed by Möller and coworkers (52) has proven to be particularly successful for smFRET-folding studies, such as those performed on RNaseH by Kuzmenkina et al. (42). However, with the protein in close proximity to a surface, a possible restriction of the conformational freedom of the polypeptide chain might bias the relationship between E_{ET} and actual interdye distance. Another elegant strategy relies on the encapsulation of one protein molecule into a vesicle of suitable dimension (diameter ≈ 120 nm $\equiv$ vol $< 10^{-18}$ liters) that is anchored to a surface via a biotin-avidin system (53). This strategy guarantees an average observation time only limited by fluorophore photobleaching (10–20 s) and at the same time ensures that the polypeptide chain is free to move in the confined space, provided that the protein does not interact with the vesicle wall (a possibility that can be largely ruled out by performing fluorescence polarization measurements). An ideal approach should allow marriage of extended observation times (as with the vesicles), with the possibility of rapidly changing the reaction conditions, as is possible in immobilized systems. In this respect, the inclusion of proteins into materials such as wet nanoporous silica gels is promising. The method, originally developed for trapping conformational intermediates of hemoglobin [(54) and references therein], has been applied to the study of the conformational dynamics of the native protein and the unfolding process of a mutant of the green fluorescent protein (GFP) both in buffer and in the presence of several concentrations of GdmCl, proving to be versatile and reliable (55, 56). Once included in the polymeric matrix (pore diameter $\approx$ 100 Å), the protein molecules can be observed for a period of hours to days (and without the energetic features of the protein being affected by confinement) (57, 58), allowing the monitoring of a complete (un)folding reaction. Conveniently, the solvent can be changed by perfusion of the desired solution through the gel, and the resulting changes monitored in real time, provided that the gel is stable in the new conditions, as seems to be the case.

Following folding trajectories of individual molecules. Despite the ability of smFRET to distinguish subpopulations in a wide range of denaturant concentrations, the structural features of the DSE at very low denaturant concentration cannot be estimated with the desired accuracy because the DSE is too scarcely populated. One way to overcome this problem is to investigate a transiently populated DSE in a FRET-monitored rapid-mixing experiment, as pioneered by Eaton and coworkers (59). Using a microfluidic mixing device to trigger the refolding of Csp*Tm* (dead time 50 ms), they highlighted a fast collapse of the polypeptide chain within the first few ms, followed by a slower compaction to the N. The results were in good agreement with ensemble stopped-flow measurements, confirming the outcome expected for a two-state protein (60, 61) and demonstrating the feasibility of coupling fast mixing methods and smFRET to gain insights into the refolding kinetics.

Rhoades et al. (53) implemented a different approach. Considering that the actual time

spent by an (un)folding molecule in crossing the barriers is, in most of the cases, too short to be detected, they concentrated their efforts in optimizing the detection of events leading from one energy well on one side of the barrier to the other (53). They used lipid vesicles to confine individual double-labeled molecules of Csp*Tm* at a concentration of denaturant corresponding to the midpoint in order to maximize the observation of (un)folding events. In this way, they were able to detect "hopping," i.e., sudden jumps between high and low E_{ET} (unfolding events) and vice versa (**Figure 3*a***). This elegant study allowed verification of the microscopic interpretation of equilibrium constants as the ratio between the kinetic constants k_f and k_u, (i.e., ratio between number of folding/unfolding events) in a protein and the discrete, stepwise, i.e., cooperative, nature of the conformational transition over the energy barrier, as expected for a two-state folder. The significance of these achievements was confirmed by a good agreement of the rate constants determined from single-molecule measurements with those obtained from ensemble experiments.

(Un)Folding pathway heterogeneity. The significant ruggedness of the energy landscape, in folding as well as in unfolding conditions, suggests that there will be a variety of trajectories by which the denatured protein can attain the native conformation. There is growing evidence for parallel pathways, but often the averaging inherent in ensemble experiments leads to underestimation of the true complexity of the folding mechanism. It is here where the true potential of single-molecule experiments should provide detailed insights into the intrinsic heterogeneity of the folding process. Indeed, by studying a vesicle-confined double-labeled domain of the adenylate kinase from *E. coli*, Haran and coworkers (53) observed a large spread of transitions (jumps between different E_{ET} values), indicative of high heterogeneity of the folding reaction, which was not detectable in ensemble

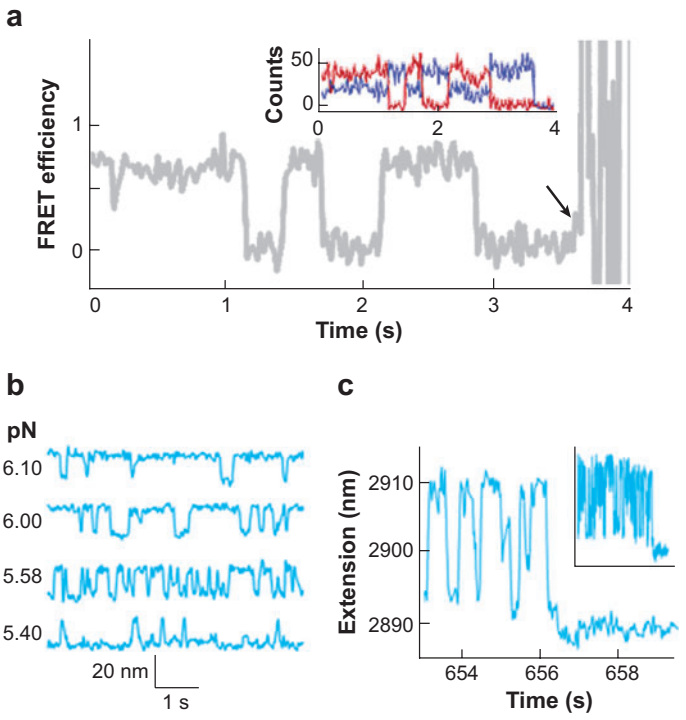


Figure 3

Single-molecule folding trajectories. (*a*) FRET efficiency trajectories of individual Csp*Tm* molecules trapped in vesicles at 2 M GdmCl showing multiple abrupt folding/unfolding transitions with a binning time of 20 ms. Arrow indicates the time of photobleaching of one of the dyes. (*Inset*) Fluorescence intensity trajectories from which FRET efficiencies were calculated are shown with the donor signal (*blue*) and acceptor signal (*red*). Panel *a* adapted and reproduced with permission from Reference 130, © 2004, American Chemical Society. (*b*,*c*) Characterization of the folding intermediate of RNaseH using optical tweezers. (*b*) Extension versus time traces for RNaseH at various constant forces. (*c*) Length versus time trace of RNaseH. Hopping stops when the protein refolds to the native state. Inset shows a longer time trace. Panels *b* and *c* adapted and reprinted from Reference 77 with permission from the American Association for the Advancement of Science.

FRET experiments previously performed by Ratner et al. (62).

Parallel unfolding pathways have also been detected for another GFP variant in a series of elegant studies (55, 56, 63). By laser illumination of individual molecules trapped in wet nanoporous silica gel over a period of hours, they were able to observe in real time the native state dynamics as well as many unfolding/refolding cycles of the same molecule. The authors detected three discrete

native conformations characterized by different oscillation frequencies for the interconversion of the anionic and neutral chemical states of the chromophore (time resolution 100 μs) (55). These different species did not interconvert, suggesting that they represent thermally inaccessible native substates. Addition of GdmCl resulted in a global unfolding event, detected by the sudden loss of fluorescence. Interestingly the time constant for unfolding was different for each native population, so that the unfolding time could be predicted from the original oscillation frequency of the chromophore in N. The results were interpreted in terms of an energy landscape where three well-defined subpopulations of N unfold through three different, parallel unfolding pathways. Upon restoration of native solvent conditions, the molecules refold to one of the native conformations in an unbiased fashion. Crucially, these substates are similar enough to evade discrimination in ensemble experiments (64).

DFS: dynamic force spectroscopy

AFM: atomic force microscopy

UNFOLDING PROTEINS BY FORCE

In these experiments, the protein is attached between two surfaces, and an external force is applied by increasing the distance between the two tethered ends. Unfolding of the protein is monitored by following the increase in length of the polypeptide chain. In dynamic force spectroscopy (DFS) experiments, the denaturant is applied force, rather than a chemical denaturant such as urea, and in principle, just as in denaturant unfolding experiments, the unfolding rate constant can be extrapolated to zero force. However, there is an important caveat, and an important difference from classical folding/unfolding experiments. When protein folding is studied at a range of urea concentrations, the explicit (and well-tested) assumptions are that folding and unfolding are the reverse of each other and that the unfolding mechanism remains the same at all denaturant concentrations. The denaturant simply acts to alter the relative stabilities of different species on the energy landscape. On the contrary, refolding in the absence of force may not be the reverse of unfolding under extreme force. The corollary of this is that force may allow proteins to explore regions of the energy landscape that are generally inaccessible in classical experiments, that is, it opens up new regions of a protein energy landscape to investigation. Furthermore, the energy landscape of many proteins under mechanical stress may be physiologically relevant.

By their very nature, forced unfolding experiments are single-molecule experiments; hence, their place in this review. Thus, in principle, forced unfolding experiments should allow us to monitor subpopulations of molecules that unfold by different pathways, perhaps, or rare unfolding intermediates. What is remarkable about the experimental studies to date is that the single-molecule aspect remains largely irrelevant; in these experiments, instead of studying many molecules at the same time, we study single molecules many times! However, there are hints that the single-molecule nature of the experiments may open up new fields for investigation.

DFS Techniques

There are currently three different DFS techniques available for studying the effect of force on proteins, atomic force microscopy (AFM), optical tweezers, and the biomembrane force probe. By far the most common is AFM—made popular by the development of a number of user-friendly off-the-shelf instruments, so that protein-folding groups have access to the technology, and consequently most of this review concentrates on AFM experiments, which are described in **Figures 4** and **5**.

Most detailed experiments have been facilitated by the use of artificial polyproteins, with a number of protein domains attached in linear array (65, 66). A single polyprotein may have a number of identical domains (typically 6–12) or different domains organized in a defined manner.

What Is Measured in DFS Experiments?

Figure 5 shows the types of traces that are obtained using AFM. The most common mode is the constant speed experiment wherein a typical sawtooth pattern of traces is observed, each corresponding to the unfolding of a single protein domain. Analysis of the traces themselves allows the length of polypeptide chain released to be determined, and thus, partly folded species to be identified, and the persistence length of the unfolded chain to be determined. The unfolding force of each domain can be measured. In analysis of the force data, the most probable unfolding force is determined at any given pulling speed, and both the unfolding rate constant at zero force (k_u) and the unfolding distance (x_u) can be estimated.

Analysis of forced unfolding data. Analysis of forced unfolding data is essentially based on the theory that force tilts the energy landscape (**Figure 1*d***) and is described only briefly here to illustrate issues that arise. Evans and coworkers (67–70) developed the theory of DFS through Kramers' work on chemical kinetics and Brownian processes in liquids (39, 40). A protein's unfolding rate, k_u, depends not only on the height of the transition state (TS) barrier and diffusion across the landscape but also on the landscape's shape. Under a persistent force, the energy landscape is tilted as $E_F(x) = E(x) - Fx$, and hence,

$$k_u(f) = \frac{D(\kappa_N \kappa_{TS})^{1/2}}{2\pi k_B T} \times \exp\left(-\frac{(G_{TS-N} - Fx_{TS-N})}{k_B T}\right) = k_u \exp\left(\frac{f x_u}{k_B T}\right), \quad 1.$$

where x_u is the distance along the force unfolding pathway from the N to the TS. There are many assumptions here. The first is the invariance in x_u as a harmonic barrier will move under force toward N, and the curvature of N will cause movement toward the TS with force, i.e. x_u is force dependent. The second is the constant exponential prefactor. This constant prefactor arises from the nature of the N and the TS, which cannot be harmonic for all locations. As the locations of the N and the TS move under force, their curvatures will deform causing a change in the prefactor. The third consideration is the asymptotic approximation of the integral that gets worse as the TS becomes nonharmonic with force.

In the normal forced unfolding experiment, the protein is stressed across a fixed

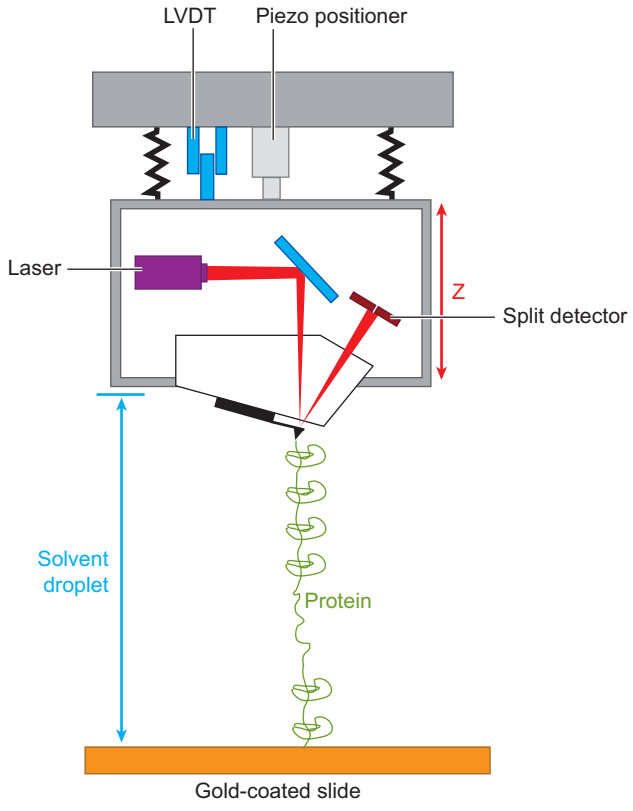
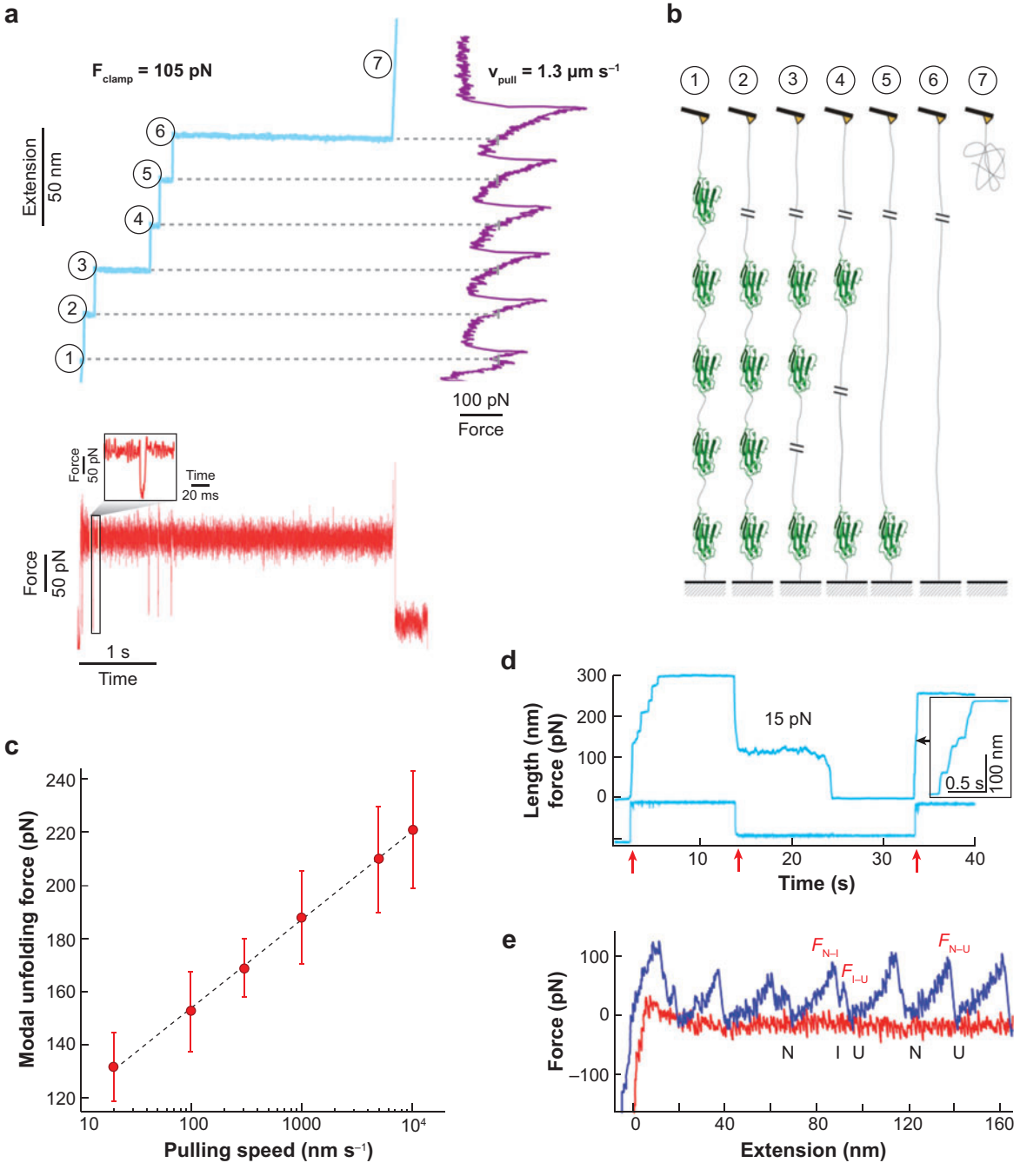


Figure 4

Schematic diagram of a force-mode atomic force microscope. The protein is attached to the surface (often through gold-sulfur bonds via Cys residues at the C terminus of the protein to a gold-coated microscope slide) and to the atomic force microscope cantilever by nonspecific adhesion to the tip. The protein is completely covered by buffer, either in a drop on the slide surface or inside a chamber. The microfabricated silicon nitride cantilever is coated with a reflective layer of gold. The cantilever deflection is detected by the deflection of a laser light onto a split photodiode. Reprinted and adapted with permission from Reference 131. Abbreviations: LVDT, linear variable differential transformer; Z, direction of separation of the tip from the sample.

TS: transition state



substrate and retracting force transducer (**Figure 4**). The histogram of unfolding forces represents a probability distribution, $p(f)$, given as the product of the unfolding probability at each force and the probability of surviving all forces to that point. When the prefactor and x_u are force independent and force is ramped slowly and constantly,

$$p(f) = \left\{ \frac{1}{r_f} k_u \exp\left(\frac{f x_u}{k_B T}\right) \right\} \times \exp\left\{ \frac{k_u x_u}{r_f k_B T} \left[1 - \exp\left(\frac{f x_u}{k_B T}\right) \right] \right\}, \quad 2.$$

and the most probable unfolding force, f^*, is

$$f^* = \frac{k_B T}{x_u} \ln\left(\frac{r_f x_u}{k_u k_B T}\right). \quad 3.$$

A plot of f^* versus $\ln(r_f)$ (a force spectrum) (**Figure 5*c***) reveals x_u and k_u. Most published AFM unfolding data fit this simple model within experimental uncertainty.

There has been considerable discussion of the presence of curvature in DFS unfolding measurements. Curvature can arise from force-induced movement of the TS and N, changes in the prefactor, changes in the landscape, and the presence of multiple TSs and pathways. Hummer & Szabo (71) derived an analysis that attributes DFS curvature to the curvature of the landscape. Reanalyzing titin I27 unfolding data, they derived $k_u \approx 10^{-4}$ sec^{-1}, x_u of 0.42 nm, and curvature at the N of 900 pNnm^{-1}, suggesting ΔG_{TS-N} between 12 and 16 kcal/mol. These conclusions are not dissimilar to others (k_u of 3×10^{-4} sec^{-1} and $x_u = 0.32$ nm) using the simple model (65, 72, 73). The data also indicate a prefactor of 10^6 sec^{-1}, agreeing with other predictions (74, 75), and suggest that DFS experiments could measure diffusion, landscape, and thermodynamic energy parameters in addition to unfolding rates and TS locations.

I27: the twenty-seventh Ig domain from the I-band of the human muscle protein titin

Schlierf & Rief (76) attributed curvature in DFS data to a broad TS. Here, an energy landscape was iterated so unfolding rates predicted by Equation 1 reproduced the measured unfolding force distributions (76). The resulting energy profile was used to evidence a funnel-shaped energy landscape. Such derivation of a potential, which reproduces the full force distribution, appears an interesting way forward for the analysis of DFS data, although

Figure 5

Sample atomic force microscopy traces. (*a*) Time course of the unfolding of single titin I27domains under constant force [upper left trace (*blue*): molecule extension, lower left trace (*red*): force]. Unfolding of each domain (*encircled numbers*) resulted in a stepwise increase of the molecular extension (*upper left trace*) and in short peaks in the force trace, (*lower left trace*). The inset shows a magnified region of the time course of the force signal detected during unfolding of a single domain. Events corresponding to unfolding I27 at constant velocity [upper right trace (*purple*)] are marked with dashed lines. The red and black traces were low-pass filtered using a cutoff frequency of 1 kHz. (*b*) Schematic representation of multidomain I27 during mechanical unfolding. Here, a cartoon with broken lines is shown; in the real experiment, the tip-surface distance increases as in (*a*) [upper left trace (*blue*)]. The numbers shown correspond to those used in (*a*). This figure was kindly supplied by Daniel Müller and was reproduced with permission from Reference 132. (*c*) Force spectrum of the unfolding of I27. The unfolding force increases with increased pulling speed. The analysis of these data allows the unfolding rate constant at 0 force (k_u) and the unfolding distance (x_u) to be evaluated. The x_u is related to the slope of the force spectrum. Any changes in pathway will result in a change in the slope. (*d*) Refolding trace; collapse of unfolded projectin domains under force. A projectin molecule was first unfolded and extended at a high force (97 pN; the applied force is shown in the lower trace), then relaxed to a force of 15 pN and extended again at a force of 97 pN. The arrows indicate jumps in the force. This figure was kindly provided by Andres Oberhauser (adapted from Reference 107). (*e*) A forced unfolding trace of a polyprotein that shows two kinds of unfolding events. The domain unfolds either by a two-state unfolding mechanism (N→U, unfolding force $F_{N\to U}$) or a three-state unfolding mechanism via an intermediate (N→I→U, unfolding forces $F_{N\to I}$ and $F_{I\to U}$), where N, I, and U are the native, intermediate, and unfolded states, respectively. Figure adapted and reprinted with permission from Reference 133.

unknown diffusion terms and their force invariance as well as their measurement and calibration errors remain serious issues.

Chevron plot: a plot of the dependence of the observed folding and unfolding rate constants on the denaturant concentration

Equilibrium Measurements

Protein unfolding usually occurs at high forces in AFM, far from equilibrium, so it is not generally possible to extract thermodynamic data from AFM unfolding experiments. However, under the low-loading-rate regime of optical tweezers, a partly folded intermediate of RNaseH has been shown to fold and refold at constant force, allowing the extraction of both forward and backward rate constants and thus determination of an equilibrium constant, between the intermediate and the unfolded state (which is the same as that determined by classical methods) (77).

Elements of secondary structure of proteins have low stability in isolation, unlike the case for short pieces of DNA. However, Bornschlogl & Rief (78) have shown that a simple coiled-coil structure will fold and unfold without hysteresis in AFM experiments, again allowing the thermodynamic parameters to be extracted from unfolding and refolding rate constants (78).

Folding and Unfolding Intermediates

In classical protein-folding experiments, two types of intermediate can be detected: There are populated, "stable" folding intermediates, which are more stable than the denatured state in 0 M denaturant, and unpopulated "high-energy" intermediates. One of the principal questions is whether such intermediates are obligatory on-pathway species or off-pathway excursions on the energy landscape. In principle, this is one of the questions single-molecule experiments could resolve.

The most elegant example comes from optical tweezers experiments on RNaseH (77). In ensemble experiments, a stable folding intermediate is populated. Such an intermediate is also detected in the force experiments. At low constant force (~5 pN) unfolding/refolding transitions between the intermediate and the unfolded state can be detected as "hops" in length (**Figure 3*b***). By varying the force, the probability of folding was determined and the stability of the intermediate evaluated. The value obtained was in remarkable agreement with that obtained from ensemble measurements. A mutant, which causes the protein to revert to two-state kinetics in the ensemble measurements, resulted in the loss of the forced unfolding intermediate. Together, this suggests that at these very low forces, forced unfolding experiments probe for the normal, force-free energy landscape. One of the most interesting aspects of this study comes from analysis of these hopping trajectories, which are reminiscent of FRET trajectories (**Figure 3*a***). Occasionally, the protein folds completely. In nearly 80% of these traces, the protein clearly folded directly from I. Because it is likely that the lifetime of I will be too short to be detected sometimes, and because the U→I transition was invariably seen in the force-relaxation curves, these results provide direct evidence to suggest that the intermediate of RNaseH is an obligatory on-pathway species.

Unfolding intermediates may be difficult to detect in classical unfolding experiments. As they are formed after the rate-limiting step for unfolding, they are not populated. They may be inferred from the downward curvature in the unfolding limb of a chevron plot (79). However, Hammond behavior, i.e., movement of the TS toward N, will have the same signature (80, 81), and it may be difficult to discriminate between the two models (82).

AFM experiments may allow us to distinguish these. Intermediates, which form after the rate-determining step for folding, may be populated in an AFM experiment when they are not in classical unfolding experiments. There are cases where a force trace shows two unfolding peaks for a single domain: a high-force peak followed by a low-force peak (**Figure 5*e***) (83, 84). This denotes formation of an unfolding intermediate: N→I→U. If sufficient chain is released on unfolding of N,

the force will drop rapidly to below the force required for I to unfold, but at the same time, it cannot refold because the residual force is too high; thus, I is "trapped" in the AFM experiments.

Hammond behavior, whereby the TS moves closer to N in destabilizing conditions, would result in a gradual decrease in x_u as the loading rate increases, resulting in an upward curvature in the force spectrum (**Figure 5*c***) with increased pulling speed. The problems with detecting such curvature are (*a*) the relative noisiness of force data (compared to classical unfolding data) and (*b*) the range of force available may be insufficient to assign the curvature with confidence. (Similarly, curvature in the unfolding limbs of chevron plots may only be observed where there is a long unfolding limb.)

Comparing Folding Pathways by Two Different Methods

Both experiments and simulations have clearly demonstrated that at high forces forced unfolding pathways are not the reverse of refolding pathways in the absence of force—force alters the energy landscape (85–88). However, it has been estimated, for titin I27, that there would be a switch to the classical unfolding pathway at low forces (~40 pN), which would require very low pulling speeds inaccessible by AFM (73).

There are, however, two cases where there is some evidence that the forced unfolding pathway is the reverse of the refolding pathway. In both of these cases, the unfolding occurs at low forces. The first case is RNaseH, pulled using molecular tweezers described in detail above (77). The second case is a domain from *Dictyostelium discoideum* filamin (ddFLN), which unfolds with a double-peak characteristic of the formation of an unfolding intermediate (84). An AFM unfolding/refolding protocol was used to investigate unfolding and formation of the intermediate. If the force is released and left for ~50 ms, the domain refolds completely, and when stretched again, the same double-peak signature is observed. However, when the protein is stretched after ~10 ms, only the unfolding of the intermediate is observed in some traces (89). The unfolding of this "refolding" intermediate has the same force signature as the intermediate observed when the fully folded protein is stretched. This is taken to mean that when the protein refolds at 0 force, it populates the same intermediate as was populated during forced unfolding.

Structural Basis for Resistance to Unfolding

Simulations clearly suggest that topology is the principle determinant of resistance to force (85, 90). All β-proteins are, on the whole, more resistant to externally applied force than all α- or α/β-proteins. The direction in which force is applied can affect the unfolding force significantly (91, 92). There are a number of recent excellent reviews on this topic [e.g., (93, 94)], so we do not cover it here. This is, however, perhaps the area where most is understood; two elegant recent studies have shown that proteins can be rationally engineered to increase the unfolding force (95, 96).

The Importance of Doing Classical Control Experiments

It is important in this discussion to learn from classical protein-folding studies. Since protein engineering was introduced as a technique, much has been learned about the role played by specific side chain interactions in protein stabilization and folding. In particular, in investigating whether a mutation has any effect on the unfolding rate (or force), the effect of mutation should be normalized against the effect a mutation has on N (4). This is essential, and trivial, yet rarely do we see such control experiments performed by the AFM community (97). Similarly, in classical folding experiments, *m*-values (the dependence of the folding and unfolding rate

constants on denaturant concentration) are examined to ensure that the mutation has not significantly altered the folding pathway or the nature of the starting state. Thus, in forced unfolding experiments, comparisons of wild-type and mutant proteins are only valid where the x_u does not alter significantly upon mutation (73, 98).

The lack of such control experiments has led to misinterpretation of the results of experiments on mutant proteins (73, 99). By contrast, the groups using FRET single-molecule experiments almost invariably measure the effects their probes have on the properties of the modified proteins.

Classical protein engineering folding experiments have also shown that one has to be very careful about interpretation of the effects of nonconservative mutations. Proline mutants are apparently a favorite of AFM groups, but these do more than delete hydrogen bonds: They change hydrophobic packing, introduce bulges into strands and kinks into helices, and generally reduce the stability of the protein significantly.

Dihydrofolate reductase as a cautionary tale. A recent controversy may highlight the importance of combining classical biophysical data with AFM experiments. When a ligand binds to a protein, the interactions formed may well alter the stability of the native or transition unfolding states of the protein (100–102), and thus, ligand binding may modulate mechanical stability. Two independent AFM studies of the effect of methotrexate on the mechanical stability of dihydrofolate reductase (DHFR) were published in the same volume of the *Biophysical Journal* in 2005.

Ainavarapu et al. (103) investigated the mechanical stability of constructs of Chinese hamster ovary DHFR (CHO-DHFR) and titin I27 in isolation and in the presence of various ligands and cofactors, including methotrexate. No unfolding peaks for the DHFR domains could be observed in any of the traces in the absence of methotrexate. The authors concluded that the DHFR domains were unfolding at very low force. On the addition of methotrexate to the sample, unfolding traces showed a number of peaks of the correct spacing for the contour length of DHFR. Unfolding forces of 82 pN and unfolding lengths of 67 nm were found for both the $(CHO\text{-}DHFR)_8$ and $(DHFR\text{-}I27)_4$ constructs.

In contrast, Junker et al. (104) studied the effect of methotrexate with mouse DHFR. Unlike Ainavarapu, Junker studied a single DHFR protein in a construct of ddFLN (ddFLN1–2-3-DHFR-ddFLN4-5). Contour length changes on unfolding were used to ascribe unfolding peaks to the DHFR or ddFLN domains, and an identical contour length of 67 nm was found for DHFR. Interestingly, the mouse DHFR alone was found to show mechanical resistance in the AFM, producing a distribution of unfolding forces of around 55 pN, increasing only slightly to 60 pN on addition of methotrexate. In over 95% of the traces, the DHFR did not unfold until at least one ddFLN domain had.

Ainavarapu attributed the differences to the method of analysis of Junker where unfolding events of less than 30 pN were excluded (105); however, the fact that in over 95% of the tests the DHFR was found to be mechanically stronger than the easily measured ddFLN domain weighs heavily against this argument (106).

The simplest explanation of the discrepancies between these two studies is the difference in source of the proteins, although they are homologous with 96% sequence identity. It seems possible that in the CHO-DHFR case the protein is not folded in the AFM polyproteins. Binding of methotrexate might be sufficient to shift the equilibrium in favor of folded protein. Alternatively, CHO- and mouse DHFR might have very different kinetic properties. These could easily be tested through classical biophysical methods, and we would argue such tests are essential controls for AFM experiments (97).

Refolding Experiments

Through the use of a force-clamp AFM, whereby a feedback loop is employed to maintain a defined cantilever deflection, chain collapse on the release of force can be observed (107, 108). If the force is low enough, then refolding of the domains occurs, detected by reapplying an unfolding force (**Figure 5*d***). Fernandez & Li (108) followed the shortening of a chain of up to eight unfolded ubiquitin proteins. The folding trajectories were described with three or four separate phases: a rapid decrease in length on quenching of the force used to unfold the proteins associated with elastic recoil, a plateau of reasonably constant length ascribed to a conformational search to find a suitable folding ensemble, and a final contraction in length, sometimes in two parts, attributed to folding. An interesting observation was the similarity in the form of these folding trajectories irrespectively of how many proteins were present in the chain. The observation sparked much interest (109), but the interpretation proved to be controversial. The authors concluded that this was evidence of highly cooperative folding whereby all domains fold together at the end of the trace. As a reflection of the significant interest that this study raised, several alternative explanations of the data were offered. Best & Hummer (110) were able to reproduce similarly shaped folding trajectories in molecular simulations in which folding was occurring throughout the trace and suggesting the entropic fluctuations in the chain would mask stepwise folding (110–113). Sosnick (113) offered an explanation of aggregation, citing previous studies where detectable aggregation of refolding ubiquitin occurs at 2 μM concentration and the effective concentration in the AFM experiments was above the mM range. Although it is clear that the collapse of an extended polypeptide chain to a condensed state can be recorded, the lack of an independent signature of folding in a force trace is a severe limitation of force measurements. Recent experiments using single ubiquitin domains should help resolve some of these issues (114).

The optical tweezers experiments on RNaseH revealed that a refolding of a protein can generate a refolding force (77). It is unsurprising that in most AFM experiments, which are far less sensitive than optical tweezers, refolding is not observed in the traces when the force is released. However, the generation of a refolding force has also been observed with tandem ankyrin repeats by traditional AFM in a remarkable study by Marszalek and coworkers (115).

Using AFM to Investigate Energy Landscape Roughness

Topological and energetic frustration gives rise to a rugged energy landscape, and the free-energy projection along the one-dimensional forced unfolding coordinate can be rough. When this roughness is significant, the unfolding rate does not follow an Arrhenius relationship. Hyeon & Thirumalai (116, 117) showed that it should be possible to measure this ruggedness using either force-clamp or DFS measurements under several values of constant temperature. Nevo et al. (118) successfully used this method to determine an energy landscape roughness for the interaction between the small GTPase Ran and the nuclear transport receptor importin-β of 5.7 k_BT. These studies showed a reduction in the location of the unbonding TS, and the expected increase in rate with increasing temperature. Similar roughness measurements were made in force-induced unfolding of transmembrane helices of bacteriorhodopsin, where again a value of 5.7 k_BT was determined with a temperature-dependent Hammond-like reduction of the x_u and an increase in k_u (119).

An interesting study of the temperature dependence of force-induced protein unfolding was described by Schlierf & Rief (120). As described above, AFM traces of the unfolding of domain 4 from ddFLN show the presence of an intermediate indicative of a three-state unfolding mechanism. Both the forces for N to I and I to U were found to decrease with increasing temperature. Here,

Hammond effect: the movement of the TS along the reaction coordinate toward the ground (native) state

x_u was estimated by analyzing the distribution of unfolding forces measured at a single loading rate, where both distributions for N→I and I→U became narrower as the temperatures were increased. The apparent shifts in x_u toward the denatured state were significant, from 0.3 to 0.8 nm for N→I and 0.45 to 0.8 nm for I→U. Furthermore, k_u for N→I appeared to decrease with increasing temperature, indicative of an increase in the barrier height. The authors describe the temperature softening of the protein with the widening of the unfolding barrier toward the unfolded state, but also provide alternative explanations. The first considers the effects of roughness on the landscape where their data can be described by a temperature-independent barrier on a landscape with approximately 4 k_BT roughness. The second explanation is a switch in the unfolding pathway. The observation of an anti-Hammond effect in classical protein-folding studies is usually taken as an indicator of such a switch (34, 79). Considering a probable difference in heat capacity between two TSs located at different distances from N, there is a reasonable expectation that a distal TS will be affected more by temperature than a proximal one, and hence, an increase in temperature could cause a switch in unfolding. The difficulty in interpreting the data fully arises through the difficulty in fitting the temperature-dependent unfolding force distribution acquired at a single pulling speed. Hyeon & Thirumalai (116) suggest the independent measurement of the precise values of the TS through a DFS study under various temperatures. Determination of x_u from a force spectrum (**Figure 5*c***) permits an independent prediction of the unfolding force distributions, which serves as a valuable control of the data.

AFM Experiments as Single-Molecule Experiments

Most of the AFM experiments described have been simply a means of applying a different type of denaturant to proteins. As such, these experiments add to our toolbox for exploring regions of the protein energy landscape that might otherwise be inaccessible. The experiments on RNaseH and ddFLN4 have shown that we can derive direct information on the role of the folding intermediate in the refolding of the protein (in both cases it is apparently on pathway) (77, 89).

However, the power of single-molecule experiments should be the ability to observe rare events or to distinguish subpopulations of molecules or folding pathways. A few AFM experiments hint at what might be achieved.

The first, and in some ways still most elegant, example of the power of single-molecule AFM experiments to observe rare events was the observation that in repeated refolding experiments on titin I27 occasional misfolded species can be detected (121). Such misfolding events are extremely rare, accounting for fewer than 1% of the refolding events. The low frequency is to be expected; protein domains, particularly those found in tandem repeat proteins, have evolved to fold efficiently to avoid misfolding. Such low frequency misfolding events would never be observed under folding conditions in classical ensemble measurements.

An intriguing observation is that detailed analysis of the unfolding kinetics of several thousands of ubiquitin molecules using force-clamp techniques identifies an apparent deviation from two-state unfolding kinetics (122). About 80% of molecules refold with a single rate constant, but the remainder of folding events cannot be accounted for by simple single exponential kinetics. Fernandez and coworkers suggested that this revealed the presence of alternative unfolding pathways. This is interesting because carefully performed classical experiments can distinguish a subpopulation of molecules folding with different rate constants when the population is over ~5%, yet single exponentials are observed in classical unfolding experiments of ubiquitin (123). An obvious explanation,

which could account for this observation, would be that interdomain interactions were affecting the unfolding. Some of the domains will be unfolding in the presence of two folded neighbors, some with one folded neighbor, and some with two unfolded neighbors. Because domain-domain interactions—both specific and nonspecific—are known to affect unfolding rate constants (124), this would be a trivial explanation. (Note that no classical folding studies, such as those we advocate above, were performed on the ubiquitin polyprotein to determine whether interdomain interactions were affecting the stability of the domains in the polyprotein.) Recently, however, by engineering longer unstructured linkers at the ends of single domains, Fernandez and coworkers (114) have reproduced the same kinetic signature from force-clamp experiments on single domain proteins. Thus, the discrepancy between classical and force-clamp AFM experiments remains. Whether the multiple exponential lifetime distribution is a property of the protein or a consequence of the measurement, such as the finite response time of the force feedback, or revealing limitations in the approximations used in the Kramers' analysis (see above) is yet unknown. Because we know that application of a force alters the unfolding landscape, it may be that this opens alternative unfolding pathways for ubiquitin. In the physiological, zero-force situation, however, a single pathway is predominant.

CONCLUSIONS

In many respects, single-molecule studies of protein folding are in their infancy, yet in this review, we have been able to discuss only a small fraction of the experiments that have been done and have not cited many interesting experimental results [we have not, for example, mentioned single-molecule fluorescence correlation spectroscopy studies (125–128)]. We have not emphasized the relationship between experiment and simulation, which is playing an important role in driving the field forward, particularly in forced unfolding work.

What we have attempted to do is to impart a flavor of the field as it stands, in particular where direct comparisons can be drawn between what has been learned in classical studies and what has been achieved using single-molecule methods. We emphasize throughout the importance of allowing classical experiments to inform both how we do single-molecule experiments and how we interpret them.

On reflection, we find that, as yet, single-molecule studies of protein folding have not provided many startling insights into the folding process and folding energy landscapes. However, a few landmark studies suggest that they have significant potential, in particular where they allow investigation of aspects of the energy landscape that are inaccessible to classical ensemble experiments.

SUMMARY POINTS

1. Single-molecule FRET techniques have allowed equilibrium populations of protein molecules to be investigated in detail. These studies demonstrate that the denatured state observed under folding conditions, at low denaturant concentrations, is significantly more collapsed than at higher denaturant concentrations. Furthermore, studies of the solvent-dependent chain contraction of Csp*Tm* suggest that the collapse of this protein is a nonspecific process, evenly distributed across the polypeptide chain, rather than driven by the early formation of specific clusters of interactions.
2. Observation of polypeptide dynamics at equilibrium allowed experimentally based calculation of the magnitude of an energy barrier separating two states that lie in consecutive energy wells, providing a new benchmark for theoretical calculations of free-energy surfaces.

3. Novel trapping techniques have allowed single molecules to be observed over significant lengths of times, so that individual folding trajectories can be monitored. Folding is demonstrably a highly cooperative process. Although some molecules appear to fold in a two-state manner, alternative (un)folding pathways and intermediate species can be observed.
4. DFS experiments (using both AFM and optical tweezers), which monitor the unfolding of a protein on application of a force, have allowed the folding pathways of a number of proteins to be investigated. This is particularly powerful when combined with simulations. Partly folded intermediates have been observed in a number of systems. A particularly elegant set of experiments using optical tweezers has allowed the unfolding and refolding of an intermediate of single RNaseH molecules to be followed in real time.
5. Single-molecule AFM experiments have shown that most molecules unfold at high forces, by pathways that are different from those observed in classical unfolding experiments, suggesting that force alters the energy landscape. This indicates that forced unfolding experiments offer the opportunity to explore regions of the protein-folding energy landscape that are inaccessible by other methods.
6. There are a number of controversial issues in the analysis and interpretation of AFM data. We note that FRET investigators make good use of classical folding techniques to validate their experiments; the same is not the case for the AFM community. We suggest that this could improve the confidence that one has in interpreting the results.
7. There have been few studies that exploit the single-molecule aspect of force-induced protein (un)folding. In the optical tweezer experiments on RNaseH, analyses of single-molecule unfolding-folding trajectories have suggested that the observed intermediate is an obligatory, on-pathway intermediate. Rare misfolding events, which have not been detected by ensemble measurements, can be detected in AFM refolding experiments.

FUTURE ISSUES

1. The development of new technologies is particularly important in this area. This development includes instrumentation with better sampling timescales and better space resolution. In the FRET field, availability of multiparameter fluorescence detection setups will become increasingly important. New "off-the-shelf" AFM instrumentation, which is easy to use, will facilitate its application by protein-folding groups. The possibility of coupling fluorescent detection and mechanical manipulation is an interesting prospect.
2. The number of new, better trapping strategies, which are being developed to allow the physical chemical conditions to be changed quickly and easily, should soon translate into the capability to monitor single molecules for long periods so that long folding trajectories can be monitored by FRET (assuming that new long-lived dyes can be developed in parallel). A photo-triggered folding reaction of a confined protein may provide an additional method for undertaking such experiments.

3. Development of new smaller and brighter stable dyes, suitable for use as a substitute for natural amino acids in protein synthesis, will allow one to progressively tailor the fluorescence properties of proteins according to more and more refined experiments (e.g., multicolor smFRET). Protein synthesis technology could soon provide a relatively accessible methodology for incorporation of these dyes directly into the polypeptide chain, reducing the time bottleneck currently represented by the labeling procedure.
4. The synergy of the above-listed technical advances will allow us to move a step toward the ultimate goal of the smFRET experiments, which is to observe step by step the relative movements of different segments of the polypeptide chain in 3D space with quasi-atomistic detail, from D to N and vice versa, in different conditions and without compromising the intrinsic kinetic and thermodynamic properties of the protein chain.
5. Up to now the single moleculeness of AFM experiments has not been exploited to any great extent. However, there have been suggestions that pathway heterogeneity can be observed. This is an exciting development and suggests that this area of the research might prove fruitful.
6. It will be particularly important to see more experiments specifically designed to address some current issues in the field of protein folding. As an example, theoretical studies suggest that single-molecule AFM experiments, performed at different temperatures, might allow exploration of the roughness of protein-folding energy landscapes. Early experiments suggest that this is indeed the case.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Ben Schuler, Daniel Müller, and Andres Oberhauser for generously supplying figures for this manuscript. We are particularly grateful to Ben Schuler and Carlos Bertoncini for their critical reading of the manuscript, in particular the fluorescence section, as none of us have any expertise in that area. Jane Clarke is a Wellcome Trust Senior Research Fellow. Alessandro Borgia is supported by a fellowship from Istituto Pasteur-Fondazione Cenci Bolognetti, Rome.

LITERATURE CITED

1. Pokala N, Handel TM. 2001. *J. Struct. Biol.* 134:269–81
2. Dill KA, Chan HS. 1997. *Nat. Struct. Biol.* 4:10–19
3. Onuchic JN, Luthey-Schulten Z, Wolynes PG. 1997. *Annu. Rev. Phys. Chem.* 48:545–600
4. Fersht AR, Matouschek A, Serrano L. 1992. *J. Mol. Biol.* 224:771–82
5. Dyson HJ, Wright PE. 2004. *Chem. Rev.* 104:3607–22
6. Scheraga HA, Khalili M, Liwo A. 2007. *Annu. Rev. Phys. Chem.* 58:57–83
7. Kelly SM, Jess TJ, Price NC. 2005. *Biochim. Biophys. Acta* 1751:119–39

8. Kumar TKS, Yu C. 2004. *Acc. Chem. Res.* 37:929–36
9. Lakowicz JR. 1999. *Principles of Fluorescence Spectroscopy.* Springer: New York. 2nd ed.
10. Förster T. 1948. *Ann. Phys.* 2:55–75
11. dos Remedios CG, Moens PD. 1995. *J. Struct. Biol.* 115:175–85
12. Haas E. 2005. *Chemphyschem* 6:858–70
13. Schuler B. 2005. *Chemphyschem* 6:1206–20
14. Deniz AA, Mukhopadhyay S, Lemke EA. 2008. *J. R. Soc. Interface* 5:15–45
15. Nienhaus GU. 2006. *Macromol. Biosci.* 6:907–22
16. Jia Y, Talaga DS, Lau WL, Lu HSM, DeGrado W, Hochstrasser RM. 1999. *Chem. Phys.* 247:69–83
17. Talaga DS, Lau WL, Roder H, Tang J, Jia Y, et al. 2000. *Proc. Natl. Acad. Sci. USA* 97:13021–26
18. Deniz AA, Laurence TA, Beligere GS, Dahan M, Martin AB, et al. 2000. *Proc. Natl. Acad. Sci. USA* 97:5179–84
19. Schuler B, Lipman EA, Eaton WA. 2002. *Nature* 419:743–47
20. Merchant KA, Best RB, Louis JM, Gopich IV, Eaton WA. 2007. *Proc. Natl. Acad. Sci. USA* 104:1528–33
21. Huang F, Sato S, Sharpe TD, Ying L, Fersht AR. 2007. *Proc. Natl. Acad. Sci. USA* 104:123–27
22. Naganathan AN, Sanchez-Ruiz JM, Munoz V. 2005. *J. Am. Chem. Soc.* 127:17970–71
23. Kuzmenkina EV, Heyes CD, Nienhaus GU. 2006. *J. Mol. Biol.* 357:313–24
24. Sherman E, Haran G. 2006. *Proc. Natl. Acad. Sci. USA* 103:11539–43
25. Hoffmann A, Kane A, Nettels D, Hertzog DE, Baumgartel P, et al. 2007. *Proc. Natl. Acad. Sci. USA* 104:105–10
26. Shastry MC, Roder H. 1998. *Nat. Struct. Biol.* 5:385–92
27. Kimura T, Akiyama S, Uzawa T, Ishimori K, Morishima I, et al. 2005. *J. Mol. Biol.* 350:349–62
28. Kimura T, Uzawa T, Ishimori K, Morishima I, Takahashi S, et al. 2005. *Proc. Natl. Acad. Sci. USA* 102:2748–53
29. Uzawa T, Akiyama S, Kimura T, Takahashi S, Ishimori K, et al. 2004. *Proc. Natl. Acad. Sci. USA* 101:1171–76
30. Welker E, Maki K, Shastry MC, Juminaga D, Bhat R, et al. 2004. *Proc. Natl. Acad. Sci. USA* 101:17681–86
31. Bowler BE. 2007. *Mol. Biosyst.* 3:88–99
32. McCarney ER, Kohn JE, Plaxco KW. 2005. *Crit. Rev. Biochem. Mol. Biol.* 40:181–89
33. Wildegger G, Kiefhaber T. 1997. *J. Mol. Biol.* 270:294–304
34. Wright CF, Lindorff-Larsen K, Randles LG, Clarke J. 2003. *Nat. Struct. Biol.* 10:658–62
35. Gianni S, Travaglini-Allocatelli C, Cutruzzola F, Brunori M, Shastry MC, Roder H. 2003. *J. Mol. Biol.* 330:1145–52
36. Borgia A, Bonivento D, Travaglini-Allocatelli C, Di Matteo A, Brunori M. 2006. *J. Biol. Chem.* 281:9331–36
37. Butler JS, Loh SN. 2007. *Biochemistry* 46:2630–39
38. Schuler B, Lipman EA, Steinbach PJ, Kumke M, Eaton WA. 2005. *Proc. Natl. Acad. Sci. USA* 102:2754–59
39. Hanggi P, Talkner P, Borkovec M. 1990. *Rev. Mod. Phys.* 62:251–341
40. Kramers HA. 1940. *Physica* 7:284–304
41. Socci ND, Onuchic JN, Wolynes PG. 1996. *J. Chem. Phys.* 104:5860–68
42. Kuzmenkina EV, Heyes CD, Nienhaus GU. 2005. *Proc. Natl. Acad. Sci. USA* 102:15471–76

43. Kubelka J, Hofrichter J, Eaton WA. 2004. *Curr. Opin. Struct. Biol.* 14:76–88
44. Nettels D, Gopich IV, Hoffmann A, Schuler B. 2007. *Proc. Natl. Acad. Sci. USA* 104:2655–60
45. Chattopadhyay K, Saffarian S, Elson EL, Frieden C. 2005. *Biophys. J.* 88:1413–22
46. Maki K, Cheng H, Dolgikh DA, Shastry MC, Roder H. 2004. *J. Mol. Biol.* 338:383–400
47. Cohen AE, Moerner WE. 2006. *Proc. Natl. Acad. Sci. USA* 103:4362–65
48. Kinoshita M, Kamagata K, Maeda A, Goto Y, Komatsuzaki T, Takahashi S. 2007. *Proc. Natl. Acad. Sci. USA* 104:10453–58
49. Rasnik I, McKinney SA, Ha T. 2005. *Acc. Chem. Res.* 38:542–48
50. Reiner JE, Crawford AM, Kishore RB, Goldner LS, Helmerson K, Gilson MK. 2006. *Appl. Phys. Lett.* 89:Artic. 013904
51. Yan M, Ge J, Liu Z, Ouyang P. 2006. *J. Am. Chem. Soc.* 128:11008–9
52. Groll J, Amirgoulova EV, Ameringer T, Heyes CD, Rocker C, et al. 2004. *J. Am. Chem. Soc.* 126:4234–39
53. Rhoades E, Gussakovsky E, Haran G. 2003. *Proc. Natl. Acad. Sci. USA* 100:3197–202
54. Viappiani C, Bettati S, Bruno S, Ronda L, Abbruzzetti S, et al. 2004. *Proc. Natl. Acad. Sci. USA* 101:14414–19
55. Baldini G, Cannone F, Chirico G. 2005. *Science* 309:1096–100
56. Cannone F, Bologna S, Campanini B, Diaspro A, Bettati S, et al. 2005. *Biophys. J.* 89:2033–45
57. Cannone F, Collini M, Chirico G, Baldini G, Bettati S, et al. 2007. *Eur. Biophys. J.* 36:795–803
58. Chirico G, Cannone F, Beretta S, Diaspro A, Campanini B, et al. 2002. *Protein Sci.* 11:1152–61
59. Lipman EA, Schuler B, Bakajin O, Eaton WA. 2003. *Science* 301:1233–35
60. Perl D, Welker C, Schindler T, Schroder K, Marahiel MA, et al. 1998. *Nat. Struct. Biol.* 5:229–35
61. Wassenberg D, Welker C, Jaenicke R. 1999. *J. Mol. Biol.* 289:187–93
62. Ratner V, Amir D, Kahana E, Haas E. 2005. *J. Mol. Biol.* 352:683–99
63. Baldini G, Cannone F, Chirico G, Collini M, Campanini B, et al. 2007. *Biophys. J.* 92:1724–31
64. Campanini B, Bologna S, Cannone F, Chirico G, Mozzarelli A, Bettati S. 2005. *Protein Sci.* 14:1125–33
65. Carrion-Vazquez M, Oberhauser AF, Fowler SB, Marszalek PE, Broedel SE, et al. 1999. *Proc. Natl. Acad. Sci. USA* 96:3694–99
66. Steward A, Toca-Herrera JL, Clarke J. 2002. *Protein Sci.* 11:2179–83
67. Evans E. 2001. *Annu. Rev. Biophys. Biomol. Struct.* 30:105–28
68. Evans E, Ritchie K. 1997. *Biophys. J.* 72:1541–55
69. Evans E, Williams PM. 2002. See Ref. 134, pp. 145–86
70. Williams PM, Evans E. 2002. See Ref. 134, pp. 187–204
71. Hummer G, Szabo A. 2003. *Biophys. J.* 85:5–15
72. Dudko OK, Hummer G, Szabo A. 2006. *Phys. Rev. Lett.* 96:108101
73. Williams PM, Fowler SB, Best RB, Toca-Herrera JL, Scott KA, et al. 2003. *Nature* 422:446–9
74. Hagen SJ, Hofrichter J, Szabo A, Eaton WA. 1996. *Proc. Natl. Acad. Sci. USA* 93:11615–7
75. Bieri O, Wirz J, Hellrung B, Schutkowski M, Drewello M, Kiefhaber T. 1999. *Proc. Natl. Acad. Sci. USA* 96:9597–601
76. Schlierf M, Rief M. 2006. *Biophys. J.* 90:L33–5

77. Cecconi C, Shank EA, Bustamante C, Marqusee S. 2005. *Science* 309:2057–60
78. Bornschlogl T, Rief M. 2006. *Phys. Rev. Lett.* 96:118102
79. Sanchez IE, Kiefhaber T. 2003. *J. Mol. Biol.* 327:867–84
80. Matouschek A, Otzen DE, Itzhaki LS, Jackson SE, Fersht AR. 1995. *Biochemistry* 34:13656–62
81. Ternstrom T, Mayor U, Akke M, Oliveberg M. 1999. *Proc. Natl. Acad. Sci. USA* 96:14854–59
82. Scott KA, Clarke J. 2005. *Protein Sci.* 14:1617–29
83. Li L, Huang HH-L, Badilla-Fernandez CL, Fernandez JM. 2005. *J. Mol. Biol.* 345:817–26
84. Schwaiger I, Kardinal A, Schleicher M, Noegel AA, Rief M. 2004. *Nat. Struct. Mol. Biol.* 11:81–85
85. Klimov D, Thirumalai D. 2000. *Proc. Natl. Acad. Sci. USA* 97:7254–59
86. Paci E, Karplus M. 2000. *Proc. Natl. Acad Sci. USA* 97:6521–26
87. Best RB, Fowler SB, Toca-Herrera JL, Steward A, Paci E, Clarke J. 2003. *J. Mol. Biol.* 330:867–77
88. Ng SP, Rounsevell RWS, Steward A, Geierhaas CD, Williams PM, et al. 2005. *J. Mol. Biol.* 350:776–89
89. Schwaiger I, Schleicher M, Noegel AA, Rief M. 2005. *EMBO Rep.* 6:46–51
90. West DK, Brockwell DJ, Olmsted PD, Radford SE, Paci E. 2006. *Biophys. J.* 90:287–97
91. Brockwell DJ, Paci E, Zinober RC, Beddard GS, Olmsted PD, et al. 2003. *Nat. Struct. Biol.* 10:731–37
92. Carrion-Vazquez M, Li HB, Lu H, Marszalek PE, Oberhauser AF, Fernandez JM. 2003. *Nat. Struct. Biol.* 10:738–43
93. Brockwell DJ. 2007. *Curr. Nanoscience* 3:3–15
94. Forman JR, Clarke J. 2007. *Curr. Opin. Struct. Biol.* 17:58–66
95. Ng SP, Billings KS, Ohashi T, Allen MD, Best RB, et al. 2007. *Proc. Natl. Acad. Sci. USA* 104:9633–37
96. Sharma D, Perisic O, Peng Q, Lam C, Lu H, Li HB. 2007. *Proc. Natl. Acad. Sci. USA* 104:9278–83
97. Rounsevell RWS, Steward A, Clarke J. 2005. *Biophys. J.* 88:2022–29
98. Best RB, Fowler SB, Toca-Herrera JL, Clarke J. 2002. *Proc. Natl. Acad. Sci. USA* 99:12143–48
99. Li HB, Carrion-Vazquez M, Oberhauser AF, Marszalek PE, Fernandez JM. 2000. *Nat. Struct. Biol.* 7:1117–20
100. Bullock AN, Fersht AR. 2001. *Nat. Rev. Cancer* 1:68–76
101. Ivarsson Y, Travaglini-Allocatelli C, Jemth P, Malatesta F, Brunori M, Gianni S. 2007. *J. Biol. Chem.* 282:8568–72
102. Sanz JM, Fersht AR. 1994. *FEBS Lett.* 344:216–20
103. Ainavarapu SKR, Li L, Badilla CL, Fernandez JM. 2005. *Biophys. J.* 89:3337–44
104. Junker JP, Hell K, Schlierf M, Neupert W, Rief M. 2005. *Biophys. J.* 89:L46–48
105. Ainavarapu SKR, Li L, Fernandez JM. 2006. *Biophys. J.* 91:2009–10
106. Rief M, Junker JP, Schlierf M, Hell K, Neupert W. 2006. *Biophys. J.* 91:2011–12
107. Bullard B, Garcia T, Benes V, Leake MC, Linke WA, Oberhauser AF. 2006. *Proc. Natl. Acad. Sci. USA* 103:4451–56
108. Fernandez JM, Li HB. 2004. *Science* 303:1674–78
109. Li MS, Hu CK, Klimov DK, Thirumalai D. 2006. *Proc. Natl. Acad. Sci. USA* 103:93–98
110. Best RB, Hummer G. 2005. *Science* 308:498
111. Brujic J, Fernandez JM. 2005. *Science* 308:498c

112. Fernandez JM, Li HB, Brujic J. 2004. *Science* 306:411c
113. Sosnick TR. 2004. *Science* 306:411b
114. Garcia-Manyes S, Brujic J, Badilla CL, Fernandez JM. 2007. *Biophys. J.* 93:2436–46
115. Lee G, Abdi K, Michaely P, Bennett V, Marszalek PE. 2006. *Nature* 440:246–49
116. Hyeon C, Thirumalai D. 2003. *Proc. Natl. Acad. Sci. USA* 100:10249–53
117. Hyeon C, Thirumalai D. 2007. *J. Phys. Condens. Matter* 19:Artic. 113101
118. Nevo R, Brumfield V, Kapon R, Hinterdorfer P, Reich Z. 2005. *EMBO Rep.* 6:428–86
119. Janovjak H, Knaus H, Muller DJ. 2007. *J. Am. Chem. Soc.* 129:246–47
120. Schlierf M, Rief M. 2005. *J. Mol. Biol.* 354:479–503
121. Oberhauser AF, Marszalek PE, Carrion-Vazquez M, Fernandez JM. 1999. *Nat. Struct. Biol.* 6:1025–28
122. Brujic J, Hermans RIZ, Garcia-Manyes S, Walther KA, Fernandez JM. 2007. *Biophys. J.* 92:2896–903
123. Krantz BB, Sosnick TR. 2000. *Biochemistry* 39:11696–701
124. Batey S, Scott KA, Clarke J. 2006. *Biophys. J.* 90:2120–30
125. Haustein E, Schwille P. 2007. *Annu. Rev. Biophys. Biomol. Struct.* 36:151–69
126. Chattopadhyay K, Elson EL, Frieden C. 2005. *Proc. Natl. Acad. Sci. USA* 102:2385–89
127. Chattopadhyay K, Saffarian S, Elson EL, Frieden C. 2002. *Proc. Natl. Acad. Sci. USA* 99:14171–76
128. Neuweiler H, Doose S, Sauer M. 2005. *Proc. Natl. Acad. Sci. USA* 102:16650–55
129. Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND. 1995. *Proc. Natl. Acad. Sci. USA* 92:3626–30
130. Rhoades E, Cohen M, Schuler B, Haran G. 2004. *J. Am. Chem. Soc.* 126:14686–87
131. Best RB, Clarke J. 2002. *Chem. Commun.* 2002(3):183–92
132. Bippes CA, Janovjak H, Kedrov A, Muller DJ. 2007. *Nanotechnology* 18:Artic. 044022
133. Ng SP, Clarke J. 2007. *J. Mol. Biol.* 371:851–54
134. Flyvbjerg H, Jülicher F, Ormos P, David F, eds. 2002. *Physics of Bio-Molecules and Cells*: *Les Houches Session LXXV*. Berlin: Springer-Verlag

Annual Review of Biochemistry

Volume 77, 2008

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Biochemistry* articles may be found at http://biochem.annualreviews.org/errata.shtml