How force unfolding differs from chemical denaturation

Guillaume Stirnemann[a], Seung-gu Kang[b], Ruhong Zhou[b,1], and Bruce J. Berne[a,1]

[a]Chemistry Department, Columbia University, New York, NY 10027; and [b]Computational Biology Center, IBM T.J. Watson Research Center, Yorktown Heights, NY 10598

Single-molecule force spectroscopies are remarkable tools for studying protein folding and unfolding, but force unfolding explores protein configurations that are potentially very different from the ones traditionally explored in chemical or thermal denaturation. Understanding these differences is crucial because such configurations serve as starting points of folding studies, and thus can affect both the folding mechanism and the kinetics. Here we provide a detailed comparison of both chemically induced and force-induced unfolded state ensembles of ubiquitin based on extensive, all-atom simulations of the protein either extended by force or denatured by urea. As expected, the respective unfolded states are very different on a macromolecular scale, being fully extended under force with no contacts and partially extended in urea with many nonnative contacts. The amount of residual secondary structure also differs: A significant population of α-helices is found in chemically denatured configurations but such helices are absent under force, except at the lowest applied force of 30 pN where short helices form transiently. We see that typical-size helices are unstable above this force, and β-sheets cannot form. More surprisingly, we observe striking differences in the backbone dihedral angle distributions for the protein unfolded under force and the one unfolded by denaturant. A simple model based on the dialanine peptide is shown to not only provide an explanation for these striking differences but also illustrates how the force dependence of the protein dihedral angle distributions give rise to the worm-like chain behavior of the chain upon force.

molecular dynamics simulations | atomic force microscopy | worm-like chain model

The protein folding problem, which has attracted a great deal of attention for the last 50 y, has greatly benefited from the interaction and the contribution of experiments, simulations, and theoretical approaches (1). Traditional, ensemble-averaged experimental techniques such as NMR, ϕ-value analysis, small-angle X-ray scattering, and many other spectroscopic techniques have provided key insights into the structural, thermodynamic, and kinetic aspects of protein folding (1, 2). More recently, advances both in time-resolved single-molecule techniques (3) and in computer simulations (1, 4) have considerably enriched our comprehension of the problem with unprecedented mechanistic details.

A key player in the protein folding event is the unfolded state, which is the initial configuration of the protein before it folds. Because the folded configuration is often the most stable monomeric protein state under physiological conditions, generation of unfolded configurations requires the application of an external perturbation able to denature the protein. In general, this is achieved by raising the temperature (thermal denaturation), changing the pH, or by adding chemical denaturants such as urea (chemical denaturation). The protein unfolded state can then be studied at equilibrium (i.e., by maintaining the denaturing conditions) by most experimental techniques (2), but can also serve as the starting configuration for protein folding after the system has been very quickly brought to native conditions by temperature quench or dilution (5). A given technique can obviously provide useful information only when its time resolution is much

higher than the characteristic folding time, as in the case of single-molecule fluorescence spectroscopy, for example (3, 5).

Another strategy to obtain unfolded proteins is to apply force using, e.g., atomic force microscopes (AFMs) or magnetic or optical tweezers (3, 6, 7). Although this approach has proved useful in probing protein folding and unfolding at a single-molecule level, the unfolded configurations generated by such techniques can be very different from the chemically unfolded ones. As an illustration of these large differences, under chemical denaturing conditions the peptide chain usually follows Flory's scaling law for a random coil in a good solvent (8, 9), whereas the force-unfolded configurations follow the worm-like chain (WLC) model of polymer entropic elasticity (10–12). In these two types of experiments, protein refolding therefore occurs from two different regions of the free-energy surface, which may influence the folding mechanism and its kinetics. Whereas previous efforts often focused on the similarity between chemical- and force unfolding (7), demonstrating most importantly that the unfolding rate under force extrapolated to zero force matches the unfolding rate in the absence of mechanical force, it is of crucial importance to better understand the differences between the unfolded state ensemble in both scenarios, which is one of the aspects we focus on here. Indeed, beyond their obvious distinct macromolecular morphologies (random coil versus extended state), little is known about more local aspects of their structure and conformations.

The structure of the chemically unfolded state by itself is a much debated topic. Small-angle X-ray scattering (SAXS) measurements suggest that the protein radii of gyration in denaturing conditions scale as $N^{0.59}$ (N is the number of residues), in agreement with the predictions from Flory's model of a random coil in a good solvent (8, 9). On the other hand, circular dichroism and NMR experiments have suggested that the unfolded state contains a significant proportion of secondary structure and native topology (13, 14), a feature which is a priori not

Significance

Proteins can be unfolded either chemically by denaturing agents (such as urea) or by mechanical force (such as optical tweezers). If these two methods produce quite different unfolded structures, the course of the subsequent protein refolding when these agents are removed will be quite different. Our computer simulations show that the respective unfolded states are indeed very different, being fully extended under force, with no contacts, and partially extended in urea with many nonnative contacts. Understanding these differences is crucial because such configurations serve as starting points for different single-molecule experimental studies of protein folding. These differences can affect both the folding mechanism and the kinetics observed, leading to very different interpretations.
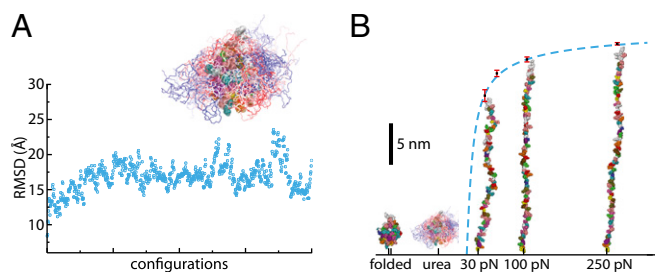
compatible with the observed random-coil behavior. However, it has been suggested that the same scaling law can actually be recovered even if the chain contains a large proportions of native segments (15, 16). Other tentative attempts to reconcile these contrasting observations have been made based on a statistical coil model derived from backbone conformational frequencies measured in the Protein Data Bank (PDB) (17).

In this paper, we use all-atom molecular dynamics (MD) simulations in explicit solvent to provide insights on the molecular features of the unfolded state ensemble. We use simulation setups previously developed in our groups to generate chemically unfolded structures (18, 19) and force-unfolded (in the range of 30–250 pN) configurations (12, 20). We focus on ubiquitin, which is a 76-residue protein well-characterized experimentally, both by bulk techniques (14, 21, 22) and by force spectroscopies (23). We perform an extensive analysis of the protein secondary and tertiary structure content in each scenario, and compare it with available experimental data. We also present an original approach based on dihedral angle distributions and their fluctuations to explain the differences between the two unfolded state ensembles.

Results and Discussion

Generation of Unfolded Configurations. Chemical denaturation was achieved by simulating the initially folded protein in a concentrated (~ 8 M) aqueous solution of urea, following a procedure already used for other proteins such as lysozyme (18, 19) (*SI Text*). However, ubiquitin is very stable against denaturing conditions and our current simulations failed to detect its complete unfolding on a timescale of ~0.5 µs. Therefore, the temperature was also raised to a high value (550 K) to allow fast unfolding of the protein along a >100-ns trajectory. (We note that raising the temperature increases the rate of conformational sampling of the unfolded state ensemble but may also affect the conformation ensemble; however, this is the only way to unfold ubiquitin on a reasonable simulation timescale.) A total of 100 configurations with a root-mean-square displacement (RMSD) >10 Å with respect to the native state were then selected at regular intervals and each of them was subsequently relaxed in the same urea solution but to ambient temperature and pressure (300 K and 1 atm) for 12 ns, leading to an aggregate simulation time of 1.2 µs under ambient conditions (doubling the relaxation time does not change the results much—see *SI Text*). A handful of such configurations (clustered to minimize the RMSD between them) is shown in Fig. 1A. Not surprisingly, the structure of the unfolded state is not unique and is characterized by a great variety of conformations. Despite these fluctuations, the RMSD of backbone C_α with respect to the native state quickly plateaus around 20 Å. It is also seen that the first $\sim$10% of the configurations exhibit a lower than average RMSD, probably because they correspond to initial structures closer to the folded state that can quickly collapse to a more native-like conformation when temperature is quenched. This small percentage of fast refolding configurations (to lower RMSDs) might be related to recent temperature jump experiments which found that some proteins, such as the λ-repressor, can refold at the protein folding "speed limit" when the denaturant is suddenly diluted (24). This superfast folding occurs from native-like unfolded populations without encountering any significant barrier.

The protein radius of gyration is distributed around an average value of 17.3 Å (*SI Text*). The corresponding ideal chain made of 76 monomers (i.e., residues) that are 3.8 Å long has an average radius of gyration of ~13.5 Å, scaling with $N^{1/2}$. As evidenced by SAXS measurements (8, 9), the protein radius of gyration in various denaturing conditions often scales as $N^{0.59}$, behaving as a random coil in a good solvent. Whereas we also find in our simulations that the protein is clearly more extended than the corresponding ideal chain, our R_g values (17.3 Å) are smaller than the experimental bare values, estimated to be $\sim$26 Å in concentrated solutions of urea and at low pH (9, 14). Considering that the radius of gyration for the folded protein is also slightly smaller in the simulations (11.9 Å vs. 13.2 Å)—mainly

Fig. 1. Generation of force- and chemically unfolded structures. (*A*) RMSD of backbone C_α atoms for a sample of chemically unfolded conformations, and superposition of protein conformations obtained after a combination of chemical and thermal unfolding (red to blue tubes), superimposed with the protein native conformation (colored spheres). (*B*) Equilibrated extended protein conformations at various forces and force-extension plot (average, black dots; SD, red bars) showing the good agreement with the worm-like chain model (blue dashed line) on this force range ($L_c = 28.4$ nm and $p = 0.39$ nm). The protein native and chemically unfolded states are shown on the same scale for comparison.

because of hydration shell diffraction (*SI Text*), the size ratio between folded and unfolded conformations is $\sim$1.5 in the simulations and $\sim$2 in the experiments (9). This discrepancy might be a consequence of the different protocols used in the simulations and the experiments, especially the fact that the pH is different. Indeed, ubiquitin is very rich in glutamic and aspartic acids (11 residues out of 76), whose protonation states are different at low and neutral pHs. It is thus expected that the protein carries a positive charge of $\sim$+13e at pH 2 (*SI Text*), which could lead to mutual repulsion between residues and to a chain extension much less pronounced than that at neutral pH. Indeed, experiments suggest that the spatial extent of the unfolded state varies greatly with the nature of the denaturant and the presence of salts, highlighting the crucial role of repulsive electrostatic interactions (25).

Force-extended configurations of ubiquitin were generated following a protocol described and used previously (12). Briefly, the protein is first unfolded at a very high force. The force is then quenched to its target value, causing the protein end-to-end distance to shrink. At each force, equilibrium properties are averaged over tens of nanoseconds once plateauing of the end-to-end distance is observed. Examples of equilibrated protein structures at different forces are given in Fig. 1B. These snapshots of the protein at representative forces suggest that in general no secondary structure is seemingly formed, despite the fact that the protein does exhibit some lateral fluctuations and local structure. The average end-to-end distances L obtained at different forces down to 30 pN, are actually well-fitted by a WLC chain force-extension profile with a contour length $L_c = 28.4$ nm and a persistence length $p = 0.39$ nm (12). This persistence length value is in remarkable agreement with that obtained by AFM experiments, including some on the same protein (7, 10, 26), suggesting that simulations reach equilibrium. However, in the present work we limit ourselves to forces above 30 pN. At lower forces (typically 10–20 pN) the barrier to collapse to a more compact and stable state is low enough that hopping will be observed in the experiment on the timescale of seconds (27). Corresponding protein simulations might encounter difficulties with the protein trapped in metastable states, even on a microsecond timescale.
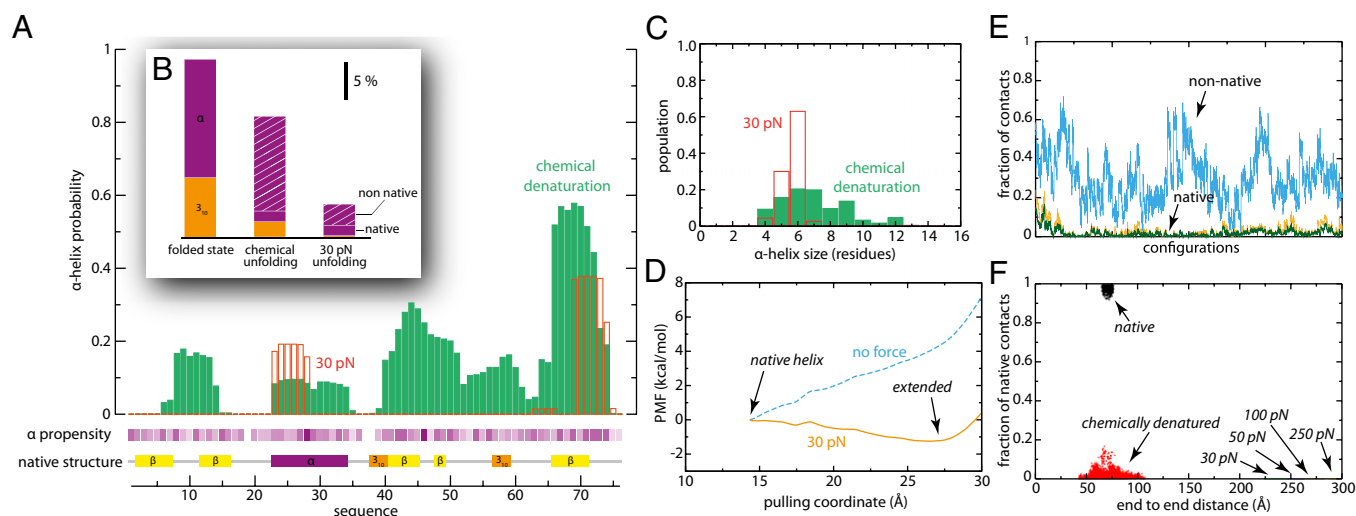
Protein Secondary and Tertiary Structure. We now characterize the long-range order observed in the protein unfolded state and discuss the differences between the two types of denaturations. There have been extensive experimental and simulation studies aimed at characterizing thermally and chemically unfolded states (8, 9, 13–16, 28). Recent breakthrough simulations have evidenced the presence of significant secondary structure, in particular helices, in the unfolded state ensemble of a large variety

of small proteins (4) [including ubiquitin (29)] in the absence of chemical denaturant, in agreement with previous experiments (28, 30). Here we make similar observations for chemically unfolded configurations (Fig. 2 *A* and *B*). As summarized in Fig. 2*B*, a significant fraction of the residues is found in a helix element (16% on average), out of which 12.6% are nonnative. Formation of the only native helix of ubiquitin (residues 23–34) is not observed in the vast majority of the structures (1.3%). The remaining 2.1% are 3–10 helices, and formation of β-sheets is not observed. These results are quite different from the secondary structure content of the ubiquitin native state (Fig. 2*B*), which contains 23.6% of helices and 31.6% of β-sheets. As shown in Fig. 2*A*, in the chemically unfolded states α-helices form almost everywhere along the sequence, whereas the folded state contains only one such helix between residues 23 and 34. These helices are relatively long, with sizes ranging from 4 to 12 residues (Fig. 2*C*). The only fragments of the proteins where helices never form contain proline residues (position 19, positions 37–38) that locally rigidify the backbone and thus cannot be accommodated inside an α-helix.

We now compare our results with previous experimental and simulations data. Whereas the formation (or nonformation) of residual secondary structure in unfolded proteins under denaturing conditions is still controversial, there has been some recent evidence for secondary structure formation based on the interpretation of NMR coupling constant and chemical shift deviation measurements in 8-M solutions of urea (30, 31). Moreover, for most proteins α-helices are found to be more resistant to denaturation compared with β-sheets (21, 31). However, experimental studies of ubiquitin denatured under high urea concentration and low pH lead to distinct conclusions, one suggesting that no secondary structure is formed at all (21) and another showing that whereas no significant populations of α-helices were observed, a native β-hairpin was present (22). As noted above, this feature is absent from our simulations performed in high concentration of urea and neutral pH, where we only observe a significant proportion of α-helices. However, the denaturing conditions in the experiments (low pH) significantly differ from ours. As already mentioned, differences in pH dramatically affect the protein charge and may have an impact on the stability of α-helices (32). This may also arise from the used force field, which is known to overestimate the stability of α-helices (33, 34).

We generally do not observe secondary structure for ubiquitin under force. The only exception is 30 pN, for which a few short α-helices form (Fig. 2*A*), mostly at one of two specific locations (although never at both simultaneously). One corresponds to a fragment of the native helix and the other is at the C-terminal end of the protein whose amino acid sequence exhibits a high helical propensity (Fig. 2*A*). Most noticeably, they are shorter than those observed in chemically denatured structures (Fig. 2*C*). Averaged over the entire trajectory, the occurrence of helical structure is 4.4% (Fig. 2*B*), which is much lower than was observed for either the chemically unfolded state (16%) or the native state (23.6%). β-sheets are never formed at any force (this would require a very improbable kink in the chain). To gain more insight into the stability of helices at the lowest force of 30 pN, we have performed distinct simulations of decaalanine in pure water, which is a model helix often used in bimolecular simulations. The native state of this polypeptide is an α-helical structure with an equilibrium end-to-end distance of 14.3 Å involving all 10 alanine residues, the amino acid with the highest helical propensity (35). As shown in *SI Text*, this helix becomes unstable under a load of 30 pN and unfolds after a few tens of nanoseconds. A more rigorous confirmation comes from the estimation of the potential of mean force (PMF) along the projection of the end-to-end distance along the pulling direction L_z using umbrella sampling (*SI Text*), which is shown in Fig. 2*D*. This PMF is in good agreement with that found earlier using a different technique and a similar force field (36). The PMF under force is simply obtained by removing the effect of force $-FL_z$ at 30 pN (being applied along z). At this force, the minimum of the PMF shifts to a higher value of ∼28 Å. This extended state is stabilized by ∼1 kcal/mol with respect to the native helical state. Because the stability of a helix correlates with its size, shorter fragments could exhibit smaller free-energy

Fig. 2. Secondary structure and contacts. (*A*) Residue probability distribution to be involved in an α-helix, averaged over the 1.2 μs of simulation for the chemically unfolded case (filled green bars) and 120 ns of simulations for the configurations at 30 pN (red bars). No α-helix is found at other forces. The α-helix propensity of each residue is also shown below, from nonprobable at all (white) to very favorable (dark purple), with a scale adapted from ref. 35 (*SI Text*). The secondary structure of the native protein is also shown. (*B*) Percentage of residues involved in secondary structure elements in the different cases (orange, 3–10 helix; plain purple, native α-helix; and hatched purple, nonnative α-helix). The fraction of π-helices is not significant in all cases, and β-strands, only present in the folded state, are not represented. (*C*) Size distribution of α-helices in the two unfolding scenarios (30 pN, red and chemical denaturation, green). (*D*) PMF of a decaalanine polypeptide in water as obtained with umbrella sampling in the absence of external force (blue), and PMF at 30 pN obtained by subtracting $-FL_z$ to the PMF in the absence of force. (*E*) Fraction of contacts in the chemically unfolded states, normalized by the number of native contacts (nonnative, blue; and native, orange). The green curve shows the reaction coordinate defined from the smoothed number of native contacts (*SI Text*). (*F*) 2D representation of the folding reaction coordinate as a function of the end-to-end distance of the protein in different conditions.

differences on the order of fraction of kT, which probably explains the transient formation of short helical fragments (accessible with thermal fluctuations) observed at 30 pN.
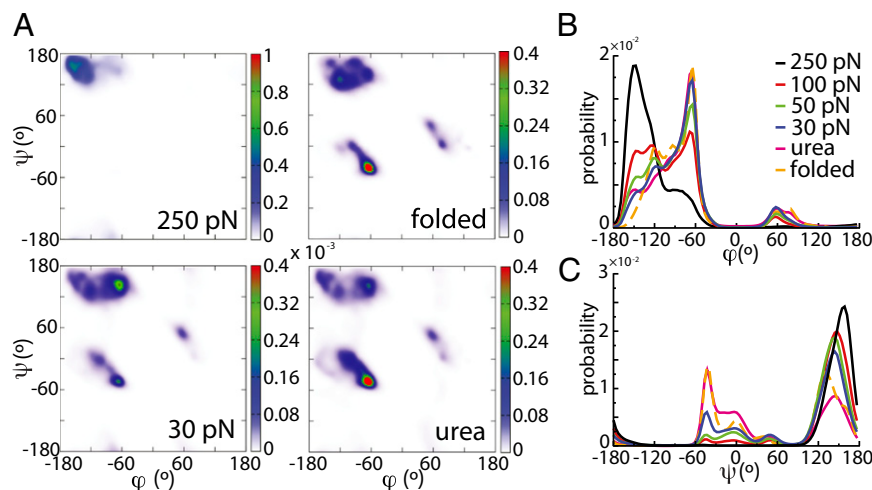
On a larger scale, protein tertiary structure is determined by the formation of key contacts between nonneighboring residues. We examine to what extent such contacts are absent or not in the unfolded state ensemble. We adopt a conventional definition detailed in *SI Text*. We make a distinction between native contacts, i.e., contacts that are present in the native folded state, and nonnative contacts, i.e., all other contacts that can form in nonnative structures but that are not present in the native state. Not surprisingly, under force no contact between nonneighboring residues is observed. As is the case with β-sheets, this would imply a serious kink in the chain that is very unlikely to happen, even at the lowest forces examined here [of course, when the forces are even smaller, such as 10–20 pN, refolding can occur (27)]. In the chemically unfolded ensemble, on the other hand, a small fraction of native contacts is observed (Fig. 2E)—on average, 3% (of fraction of contacts present in the folded state), with some configurations showing up to 20% of native contacts, which correspond to the structures with the smaller RMSD (Fig. 1A) and the formation of the native α-helix (Fig. 2A). In addition, a significant population of nonnative contacts is found in most of the configurations (average 30%, standard deviation 12%) indicating that whereas these structures are nothing like the native state, they form a molten-globule state with many contacts between nonneighboring residues and a significant amount of α-helices. Another striking observation is that the proportion of nonnative contacts fluctuates between 0% and 70% among the chemically unfolded structures, whereas variations of the RMSD (Fig. 1B) are much less pronounced. As expected, this illustrates the large variety of tertiary structures present in these unfolded configurations. Cluster analysis of MD trajectories has suggested that unfolded states can actually form several basins in the free-energy landscape (29) that sometimes interconvert more rapidly with the folded basin than between each other (37).

Force spectroscopy monitors the response of the end-to-end distance of a protein (or polyprotein) to a mechanical perturbation and the fluctuation dynamics of the end-to-end distance is limited by the object to which the protein is tethered (20). In force-quench experiments the end-to-end distance decreases as the protein refolds, yet it is a very ineffective reaction coordinate for folding from a chemically denatured state (Fig. 2F). The initial end-to-end distance in such experiments is much larger than in the chemically denatured state, but this distance is almost the same in the chemically denatured state as in the native state of the protein, albeit with a broader distribution (Fig. 2F). Thus, protein folding after force has been quenched differs significantly from folding after the chemically denatured
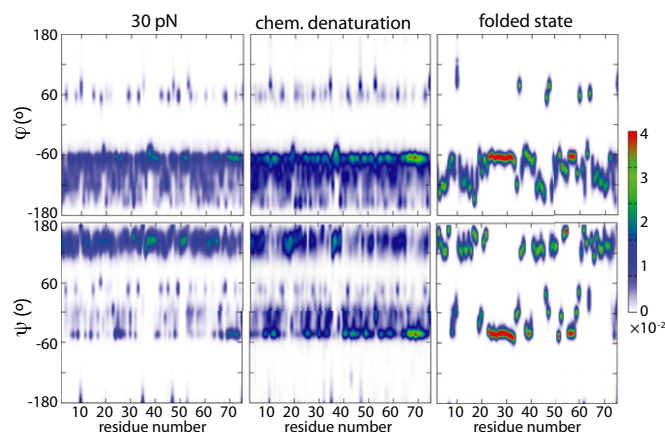
protein is quickly diluted. The two differently prepared initial ensembles will explore different regions of phase space in the early stages of folding and might well get trapped in different metastable free-energy basins. We expect that both ensembles will eventually reach similar "doorway" states before the final folded state is reached, but this we cannot say with any certainty.

Local Structure. A convenient description of the chain conformation on a residue basis is provided by the backbone dihedral angles. Whereas the $\omega = C_{\alpha(-1)}C_{(-1)}NC_\alpha$ is readily close to 180° in any circumstance as imposed by conjugation, the two other dihedral angles $\phi = C_{(-1)}NC_\alpha C$ and $\psi = NC_\alpha CN_{(+1)}$ are known to give many insights about the protein local structure (17) and the force sensitivity (12). The Ramachandran plots of the protein in various conditions are presented in Fig. 3A. As already noted (12), at high force (250 pN and above) most of the population is concentrated in the top left corner with (ϕ,ψ) values close to (−180°, 180°), corresponding to a fully extended backbone. The Ramachandran plot is radically different from that of the folded structure in the absence of force. Upon relaxation to more moderate forces (e.g., 50 pN), dihedral angles start exploring lower angle values and the backbone becomes more compact. The peak around (−180°, 180°) becomes more spread out and the populations at positions characteristic of the coil fragments of the PDB (17) grow, mainly in the so-called "polyproline II" (PPII) region $\approx (−75°, 150°)$. A small population is found in the region corresponding to α-helices at $\approx (−75°, −60°)$, even if this does not necessarily mean that helices are actually formed. The Ramachandran plot of the chemically denatured ensemble resembles to some extent the low force case (30 pN), with two main peaks in the PPII and α-helical regions. Of course, the latter is now much more populated than the random-coil region, for, as shown in the previous section, a significant amount of α-helices is found in the urea-denatured configurations.

To better describe the discrepancies between the Ramachandran plots under force and in chemical denaturant, we provided a one-dimensional projection along each axis (simply corresponding to the probability distributions of each dihedral angle), as shown in Fig. 3 B and C. ϕ is very sensitive to force (Fig. 3B): when force is lowered, the peak close to −180° progressively shifts to higher values while disappearing, and sharp peaks around −60° and 60° (having similar peptide end-to-end distance) that are absent at 250 pN rise (12). The ϕ-distribution for the chemically unfolded ensemble is quite similar to that at 30 pN. This is expected because the main difference observed in the Ramachandran was in the respective populations in the PPII and helical region, which both have similar ϕ-values but distinct ψ-values. It is also remarkable that both distributions for unfolded structures are close to that of the folded, native state, but

Fig. 3. (A) Ramachandran plots. (*Upper*) From left to right, 250 pN and folded at no force; (*Lower*) from left to right, 30 pN, chemically denatured. Note the scale differs between 250 pN and the other cases. (B) Distribution of the ϕ-dihedral angle. (C) ψ for unfolded ubiquitin at different forces and compared with the distribution for the chemically unfolded and the folded, equilibrium protein at no force.

Stirnemann et al.

Fig. 4. Distribution of dihedral angles ϕ and ψ for each residue along the protein sequence for the unfolded protein at 30 pN (*Left*), the chemically unfolded protein (*Center*), and that of the folded protein (*Right*).

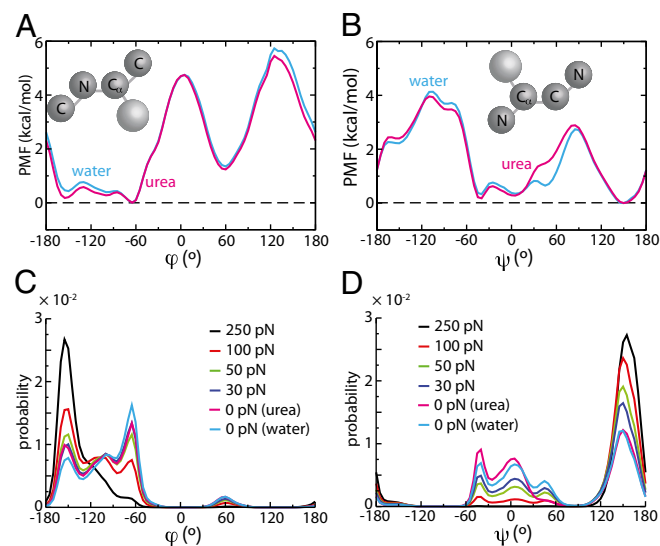these similarities hide a very different dynamical behavior, as discussed below.

The ψ-distributions in different conditions are shown in Fig. 3*C*. As force is lowered (12) the large peak near 180° at high force slightly shifts to lower ψ-values and its amplitude decreases. Concomitantly, the region of the distribution between −60° and 60°, leading to shorter peptide end-to-end distance, becomes more populated. In chemically denaturing conditions, the distribution now differs from that at a low force of 30 pN. Three main peaks around −50°, 0°, and 150° are present in both the 30-pN and chemical denaturing cases, but their respective amplitudes are different. This reflects the fact that the chemically unfolded structures exhibit a significant population of α-helices, giving rise to the sharp peaks at −50° and 0°, whereas the force-unfolded configurations are mainly random coil, with a $\psi \approx 150°$. In contrast with ϕ, none of the distributions coincides with that of the folded protein.

As discussed above, the average 1D and 2D distributions of dihedral angles at low force and under chemical denaturation may look similar to that of the folded protein, especially for ϕ. However, they are very different in nature, as illustrated in Fig. 4. For the folded state, the heterogeneous distribution arises from static heterogeneities among the amino acid sequence. The dihedral angle fluctuations of each residue are very limited because they often correspond to a particular local secondary and tertiary structure. Once averaged over the entire sequence, the distribution is broad for these different local structures corresponding to various values of dihedral angles. The situation is very different under a low force or in the presence of a chemical denaturant. Under force, where almost no secondary structure is observed, the individual fluctuations for each residue are very similar along the sequence (with the notable exceptions of the short region centered around the two consecutive proline residues 37 and 38, which make the chain more rigid locally). In this case, the average distribution is almost insensitive to the nature of the amino acid and therefore arises from dynamic disorder. The situation is similar for the chemically unfolded structure, in particular for the fluctuations of ϕ; for ψ, the averaged distributions differ because of the transient formation of α-helices, a feature even more prominent when examining Fig. 4. Indeed, for most of the residues, fluctuations around the α-helical and PPII basins are observed, with two noticeable exceptions. One corresponds to the rigid proline-rich region around residues 37–38 as mentioned above, and the other to the most stable non-native helix around residues 65–75, which as shown in Fig. 2*A* is present in more than 80% of the configurations.

To gain more insight into the system- and the force dependence of the dihedral angle distribution, we have performed

additional simulations on a model dialanine peptide to estimate the PMF along the ϕ- and ψ-dihedral angles (*SI Text*). In Fig. 5 *A* and *B*, we show the PMFs obtained in pure water and in an aqueous solution of urea. The PMFs in pure water illustrate the conformation preferences of both dihedral angles in the absence of any perturbation. Along the ϕ-dihedral angle, a low-energy region is located between −170° and −60°, whereas another local minimum is present around +60°. These regions are separated by barriers on the order of 5 kcal/mol. The distribution of the ψ-dihedral angle exhibits two free-energy energy basins between −60° and +60° and around +150°, separated by barriers on the order of 3–4 kcal/mol. Fig. 5 *A* and *B* also shows that the PMFs in pure water and in aqueous urea solution are quite similar to each other. This suggests that in this model peptide, the local backbone structure is little sensitive to the presence of the denaturant (of course, this observation cannot be easily generalized for a real protein system because other side chains may be more sensitive to urea or other denaturants, but for our current discussion it serves our purpose sufficiently).

The dihedral angle distributions corresponding to the PMFs are shown in Fig. 5 *C* and *D*. We also consider the effect of force on these distributions using a small analytical model where we estimate, for each value of a given dihedral angle, the extension δL between the first and last atoms with respect to their closest separation corresponding to a dihedral angle value of 0°. The effect of force F on the PMF is obtained by adding $-F \times \delta L$, as detailed in *SI Text*. When force is increased, distributions are obviously shifted toward higher absolute values of the dihedral angles. Quite remarkably, this simple model reproduces the observed force dependence presented in Fig. 3 almost quantitatively, in particular for the peak positions and respective amplitudes. Another observation is that the ϕ-distributions under a force of 30 pN and in the presence of urea are again very similar, whereas the corresponding ψ-distributions exhibit more noticeable differences, as was also observed in the real protein system (Fig. 3 *B* and *C*). The lower energy barrier along ψ explains why its distribution is more distorted by the external forces, particularly when the force is large. Finally, the free-energy barriers of at most 5 kcal/mol for any individual dihedral angle in the absence of force also provide an explanation for the dynamic disorder suggested above. Indeed, such barriers can be crossed on a nanosecond timescale so that for a given dihedral

Fig. 5. Predictions based on a model dialanine system. PMF along the ϕ (*A*) and ψ (*B*) dihedral angle coordinate, both in pure water and in a concentrated aqueous solution of urea. The corresponding distributions are also shown for ϕ (*C*) and ψ (*D*), together with the estimated distributions under force using a simple model detailed in *SI Text*.

angle in our simulations, exchange between the two free-energy basins occurs on this timescale, which is at least one to two orders of magnitude faster than the length of the MD trajectories.

Concluding Remarks

In this paper we examine the local and long-range structures of unfolded ubiquitin under two different denaturing conditions, namely force and chemical denaturation. We examined the backbone dihedral angle distributions, which are a good metric for the local structure, as well as the secondary and tertiary structures as characteristics of the longer range structural elements. These two ways of unfolding a protein seemingly result in very different physical descriptions: a fully extended peptidic chain whose extension with force follows the worm-like chain model (if the force is not too low) and a "molten-globule-like" structure not very much different from the native folded state in terms of spatial extension for the chemically denatured chain. Even under a low force of 30 pN, formation of secondary structure elements is not favorable and only a few short helices are observed; moreover, no contacts (native or nonnative) form between nonneighboring residues. In contrast, the chemically denatured structure contains a large proportion of (mostly) nonnative α-helices. Although these configurations do not resemble the native state as almost no native contacts are formed, a significant number of nonnative contacts are observed, as expected. More strikingly, force- and chemically unfolded proteins exhibit some differences in their local chain structure (even under relatively low forces). In particular, the distributions of ψ-dihedral angles are markedly different in the two cases. Despite a similar location of the main peaks, their respective populations in each peak differ significantly because of the balance between α-helices (mostly present in the chemically denatured states) and random coils (mostly present in the force-unfolded chain). The Ramachandran plots subsequently reflect these differences.

Another striking feature is that despite the similar average dihedral angle distributions for unfolded proteins and the folded one, a careful examination of these distributions for each residue independently shows that they are very different in nature. For the folded state, the heterogeneous distribution arises from static heterogeneities among the amino acid sequence, whereas under force or in the presence of denaturant, most of the amino acids along the sequence sample a significant part of the distribution, which is therefore due to dynamic disorder. This illustrates the fundamental difference between a folded state with fixed and well-defined local geometries and unfolded state configurations where there is almost no conformational preference among residues. Finally, we have developed a simple model based on the dialanine peptide that is able to capture and reproduce the observed force dependence of the protein dihedral angle distributions, and which provides an explanation of the dynamic disorder seen in the chemically denatured ensemble and at low force.

Materials and Methods

All-atom MD simulations of ubiquitin were performed in explicit solvent. All subsequent analysis of the trajectories was done with homemade codes as well as standard and publicly available algorithms. A full description of the simulations and the analysis is given in *SI Text*.

1. Dill KA, MacCallum JL (2012) The protein-folding problem, 50 years on. *Science* 338(6110):1042–1046.
2. Bartlett AI, Radford SE (2009) An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat Struct Mol Biol* 16(6): 582–588.
3. Borgia A, Williams PM, Clarke J (2008) Single-molecule studies of protein folding. *Annu Rev Biochem* 77:101–125.
4. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. *Science* 334(6055):517–520.
5. Lipman EA, Schuler B, Bakajin O, Eaton WA (2003) Single-molecule measurement of protein folding kinetics. *Science* 301(5637):1233–1235.
6. Neuman KC, Nagy A (2008) Single-molecule force spectroscopy: Optical tweezers, magnetic tweezers and atomic force microscopy. *Nat Methods* 5(6):491–505.
7. Carrion-Vazquez M, et al. (1999) Mechanical and chemical unfolding of a single protein: A comparison. *Proc Natl Acad Sci USA* 96(7):3694–3699.
8. Kohn JE, et al. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci USA* 101(34):12491–12496.
9. Millett IS, Doniach S, Plaxco KW (2002) Toward a taxonomy of the denatured state: Small angle scattering studies of unfolded proteins. *Adv Protein Chem* 62:241–262.
10. Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE (1997) Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* 276(5315):1109–1112.
11. Bustamante C, Marko JF, Siggia ED, Smith S (1994) Entropic elasticity of lambda-phage DNA. *Science* 265(5178):1599–1600.
12. Stirnemann G, Giganti D, Fernandez JM, Berne BJ (2013) Elasticity, structure, and relaxation of extended proteins under force. *Proc Natl Acad Sci USA* 110(10): 3847–3852.
13. Shortle D, Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. *Science* 293(5529):487–489.
14. Kamatari YO, et al. (1999) The compact and expanded denatured conformations of apomyoglobin in the methanol-water solvent. *Protein Sci* 8(4):873–882.
15. Wang Z, Plaxco KW, Makarov DE (2007) Influence of local and residual structures on the scaling behavior and dimensions of unfolded proteins. *Biopolymers* 86(4): 321–328.
16. Fitzkee NC, Rose GD (2004) Reassessing random-coil statistics in unfolded proteins. *Proc Natl Acad Sci USA* 101(34):12497–12502.
17. Jha AK, et al. (2005) Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. *Biochemistry* 44(28):9691–9702.
18. Gao M, She ZS, Zhou R (2010) Key residues that play a critical role in urea-induced lysozyme unfolding. *J Phys Chem B* 114(47):15687–15693.
19. Hua L, Zhou R, Thirumalai D, Berne BJ (2008) Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc Natl Acad Sci USA* 105(44):16928–16933.
20. Berkovich R, et al. (2012) Rate limit of protein elastic response is tether dependent. *Proc Natl Acad Sci USA* 109(36):14416–14421.
21. Wirmer J, Peti W, Schwalbe H (2006) Motional properties of unfolded ubiquitin: A model for a random coil protein. *J Biomol NMR* 35(3):175–186.
22. Meier S, Strohmeier M, Blackledge M, Grzesiek S (2007) Direct observation of dipolar couplings and hydrogen bonds across a beta-hairpin in 8 M urea. *J Am Chem Soc* 129(4):754–755.
23. Popa I, Fernández JM, Garcia-Manyes S (2011) Direct quantification of the attempt frequency determining the mechanical unfolding of ubiquitin protein. *J Biol Chem* 286(36):31072–31079.
24. Yang WY, Gruebele M (2004) Folding λ-repressor at its speed limit. *Biophys J* 87(1): 596–608.
25. Guo L, Chowdhury P, Glasscock JM, Gai F (2008) Denaturant-induced expansion and compaction of a multi-domain protein: IgG. *J Mol Biol* 384(5):1029–1036.
26. Carrion-Vazquez M, et al. (2003) The mechanical stability of ubiquitin is linkage dependent. *Nat Struct Biol* 10(9):738–743.
27. Berkovich R, Garcia-Manyes S, Klafter J, Urbakh M, Fernández JM (2010) Hopping around an entropic barrier created by force. *Biochem Biophys Res Commun* 403(1): 133–137.
28. Qin Z, Ervin J, Larios E, Gruebele M, Kihara H (2002) Formation of a compact structured ensemble without fluorescence signature early during ubiquitin folding. *J Phys Chem B* 106(50):13040–13046.
29. Piana S, Lindorff-Larsen K, Shaw DE (2013) Atomic-level description of ubiquitin folding. *Proc Natl Acad Sci USA* 110(15):5915–5920.
30. Bai Y, Chung J, Dyson HJ, Wright PE (2001) Structural and dynamic characterization of an unfolded state of poplar apo-plastocyanin formed under nondenaturing conditions. *Protein Sci* 10(5):1056–1066.
31. Schwalbe H, et al. (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry* 36(29):8977–8991.
32. Shoemaker KR, et al. (1985) Nature of the charged-group effect on the stability of the C-peptide helix. *Proc Natl Acad Sci USA* 82(8):2349–2353.
33. Best RB, Buchete NV, Hummer G (2008) Are current molecular dynamics force fields too helical? *Biophys J* 95(1):L07–L09.
34. Lindorff-Larsen K, et al. (2012) Systematic validation of protein force fields against experimental data. *PLoS ONE* 7(2):e32131.
35. Pace CN, Scholtz JM (1998) A helix propensity scale based on experimental studies of peptides and proteins. *Biophys J* 75(1):422–427.
36. Ozer G, Quirk S, Hernandez R (2012) Thermodynamics of decaalanine stretching in water obtained by adaptive steered molecular dynamics simulations. *J Chem Theory Comput* 8(11):4837–4844.
37. Bowman GR, Pande VS (2010) Protein folded states are kinetic hubs. *Proc Natl Acad Sci USA* 107(24):10890–10895.

Supporting Information

Stirnemann et al. 10.1073/pnas.1400752111

SI Text

Simulation Details

Generalities. All-atom simulations of wild-type ubiquitin [Protein Data Bank (PDB) ID code 1UBQ] in explicit solvent were carried out with the NAMD 2.8 (1) software, using the CHARMM22 force field with CMAP corrections for the protein (also known as CHARMM27) (2), and the TIP3P-CHARMM water model. Periodic boundaries conditions and a cutoff of 12 Å for electrostatic and Lennard-Jones interactions were used. Long-range electrostatic interactions were calculated using the PME method (3) with a grid spacing of 1 Å. All bonds between light and heavy atoms were maintained rigid, whereas the rest of the protein was flexible. Overall, the trajectories used in this work represent a total simulation time of ~2 μs.

Choice of Force Field. CHARMM27 (2), which includes the CMAP corrections, provides a very good description of the backbone dihedral distributions, which is one of the main focuses of the present work. Some exploratory data using the CHARMM22 force field suggested that this latter one behaves very poorly in that respect. As pointed out in numerous studies (see, e.g., ref. 4), the CHARMM27 force field (as well as many others) tends to overstabilize helices. However, as shown in ref. 5, CHARMM27 scores reasonably well in reproducing experimental NMR J coupling constants even before reweighting, and folding of villin was observed *in silico* using this force field (4). We believe it represents a good compromise which is widely used by the community for protein folding simulations.

Force Unfolding Simulations. Steered MD simulations were performed by fixing the C_α of the first residue (MET1) and by applying a constant force on the C_α of the last residue (GLY76) along the z direction. These simulations were performed on a local cluster. To unfold the protein, we first pulled on ubiquitin molecule in vacuum at a high force of 800 pN, during 10 ns. A fully extended protein was thus generated, with no remaining secondary structure. It was then solvated using the waterbox module of visual molecular dynamics (VMD) (6) in a box of 3.5 × 3.5 × 32 nm, comprising 11,499 water molecules and 35,728 atoms total. Energy minimization using the steepest descent method (2,000 steps) was performed before further equilibration, as described below. The protein was then equilibrated for 6 ns at 250 pN in the isobaric ensemble at 300 K and 1 bar, using a time step of 2 fs, a Langevin thermostat (damping coefficient of 1/ps) for temperature control, and the modified NAMD version of the Nosé–Hoover barostat with Langevin dynamics (piston period of 0.1 ps and piston decay time of 0.05 ps) for pressure control. This simulation was then propagated for 25 ns more to check that the average end-to-end distance no longer evolved. A similar procedure was applied to generate trajectories at other forces starting from a configuration at 250 pN. Once plateauing of the end-to-end distance was observed (after ~3 ns at 100 pN but ~60 ns at 30 pN), the simulations were later propagated as described above for more than 25 ns. At the lowest force studied here (30 pN), we generated two such trajectories to check that they collapsed to the same average value of the end-to-end distance. (We accumulated data for more than 120 ns at this force.) At each force the average end-to-end distance no longer evolved during these production runs, yet fluctuations were observed. All these trajectories were used for the subsequent structural analysis.

Chemical Unfolding Simulations. In the chemical denaturation simulation, we prepared a system with ubiquitin immersed in 8-M urea solution. The protein was solvated by patching preequilibrated 8-M urea solution around the protein, where the urea solution has been set up with 30 urea and 128 water molecules with a 100-ps NVT simulation at 310 K (7). Any heavy atoms of urea and/or water were removed when overlapped with protein in distances less than 2.4 and 2.7 Å, respectively, generating a system of 34,000 atoms in a box of 6.9 × 6.9 × 6.9 nm. The denatured ensemble has been obtained by two separated simulation procedures. Because ubiquitin is a highly thermostable protein (8), we first ran this simulation for more than 120 ns with a 1.5-fs time step at a temperature of 550 K to facilitate unfolding of ubiquitin (the protein was stable even at 525 K in our test), among which we selected 100 different configurations. Only those configurations with a root-mean-square displacement (RMSD) larger than 10 Å from the native fold and separated by at least 1 ns from each other were selected. Each of them was then annealed down to run for an additional 12-ns trajectory with NPT ensemble of 1 atm and 300 K, hence generating an aggregate of 1.2 μs of production run at 300 K. All simulations were performed with NAMD2 software, which was massively parallelized in IBM Blue Gene supercomputer (9). Each simulation was first energy minimized for 20,000 steps to remove undesired van der Waals overlaps from thermal denaturation simulation, followed by 250-ps equilibration with a 0.5-fs time step before leading to production run for data analyses.

Helix Stability at 30 pN. To investigate the stability of α-helices at the lowest force considered here (30 pN), we have performed additional simulations of decaalanine, which in the absence of force is purely helical, and whose length corresponds to the average α-helix length in the PDB. The system was first minimized in a box of 3.06 × 2.79 × 4.74 nm as detailed above for ubiquitin. The starting PDB structure was taken from the VMD website (6), and it has an end-to-end distance (between ALA1 C_α and ALA10 C_α) of 13.73 Å. A constant-velocity simulation of 2 ns in the NPT ensemble (same parameters as used for ubiquitin) was then performed by fixing the ALA1 C_α and pulling the ALA10 C_α at a rate of 0.01 Å/ps with a stiff harmonic potential (force constant 5 kcal·mol^{-1}·Å^{-2}). As a result, the end-to-end distance L increased almost linearly with time from its initial value and up to 34 Å where this simulation was stopped. Configurations at different end-to-end distances served as starting configurations of umbrella-sampling simulations at no force, as detailed below. We sampled the projection of the end-to-end distance along the z axis (L_z) from 14 to 30 Å with a force constant of 5 kcal·mol^{-1}·Å^{-2} and we fixed the ALA1 C_α. We constrained the molecule along the vertical z axis so that the molecule is not rotating and therefore a smaller box size in other directions can be used. At each L_z, an equilibration simulation was first performed for 2 ns before a production run was propagated for 50 ns, and used for subsequent analysis. The potential of mean force (PMF) at no force was finally reconstructed using the program WHAM (http://membrane.urmc.rochester.edu/content/wham). The PMF under force was simply recovered by removing the term $-F(L_z - L_{eq})$ to the PMF at no force (L_{eq} designates the equilibration length, i.e., the minimum at no force, and is equal to 14.3 Å).

Polymer Models. The ideal chain model considers a polymer made of N rigid rods of individual length l whose positions and orientations are random and independent of each other. The radius of gyration for such a chain is given by

$$R_G = \sqrt{\frac{N}{6}} l. \qquad \textbf{[S1]}$$

Within the worm-like chain model, the polymer is described as a semiflexible rod fully characterized by its contour length L_c (total length) and its persistence length p (which corresponds to the chain spatial "memory"). The end-to-end distance L of the polymer under force is approximated by (10)

$$\frac{Fp}{k_B T} = \frac{1}{4} \left[\frac{1}{\left(1 - \frac{L}{L_c}\right)^2} - 1 \right] + \frac{L}{L_c}. \qquad \textbf{[S2]}$$

Structural Elements. Data analysis was performed using homemade codes, except secondary-structure assignment that was done using the VMD implementation of the software STRIDE (11) on a set of regularly spaced configurations (4,000 for chemical unfolding and 5,200 at 30 pN). The α-helical propensity of each individual residue was estimated using the experimental scale of ref. 12. The energy penalty to form the helix (compared with the alanine reference that has the highest helical propensity) was translated into a probability scale simply writing that $P = \exp(-\Delta G/kT)$, with $P = 1$ for alanine and $P \approx 0$ for proline, which has the lowest helical propensity. The contacts between nonneighboring residues were defined following ref. 13. Two residues were considered to be in contact if their C_αs are closer than 10 Å apart and if they are separated by more than seven residues in the sequence. We were thus able to define the total number of contacts (native + nonnative) for each given protein configuration. The definition of native contacts involves the generation of a folded protein ensemble, obtained as follows: 4,000 protein configurations were selected along a separate 20-ns simulation of the folded protein at equilibrium. These configurations were clustered using a C_α RMSD cutoff of 2 Å, and the most populated cluster was defined as the folded state ensemble. Native contacts were then defined as contacts present more than 80% of the time in this cluster. All other contacts are designated as nonnative. A contact-based reaction coordinate was also used, following ref. 13. It is defined as

$$Q(t) = \frac{1}{N_c^{tot}} \sum_{i=1}^{n_{res}} \sum_{j=1}^{N_c(i)} \frac{1}{1 + e^{10\left[d_{ij}(t) - (d_{ij}^0 + 1)\right]}}, \qquad \textbf{[S3]}$$

where N_c^{tot} is the total number of native contacts, n_{res} is the number of residues, $N_c(i)$ the number of contacts of the ith residue, and $d_{ij}(t)$ (respectively, d_{ij}^0) the distance between the C_α of residue i and that of residue j at time t (respectively, in the crystal structure). The exponential form ensures a smoothing of an otherwise discrete quantity.

Dihedral Angle PMF. The PMF along the protein backbone dihedral angles was estimated by applying the adaptive biasing forces (ABF) method (14, 15) for the alanine dipeptide (N-acetyl-N-methylalaninamide). Simulations were performed in two different solutions: pure water and 8-M urea in a box of $2.5 \times 2.5 \times 2.5$ nm. With the same force field used in ubiquitin simulations (CHARMM27), conformations were independently sampled for each backbone torsion angle (i.e., ϕ- and ψ-angles) ranging between $-180°$ and $180°$ every $5°$. Simulations proceeded with a 1-fs time step over 20 ns in the NPT ensemble of 1 atm and 300 K. The ABF were applied only after a threshold of 100 force samples, and at least 200,000 samples were collected to calculate the PMF.

To estimate the effect of force on the dihedral angle PMF, and subsequently on the dihedral angle distributions, we first calculated the distance between the first and the last atom involved in the dihedral angle for each value between $-180°$ and $180°$. This was done analytically by assuming constant bond lengths and angles that we estimated by averaging over all backbone atoms in the equilibrium simulation of ubiquitin ($d_{C-N} = 1.346$ Å, $d_{N-C_\alpha} = 1.455$ Å, $d_{C-C_\alpha} = 1.522$ Å, $\widehat{NC_\alpha C} = 111.6°$, $\widehat{C_\alpha CN} = 117.6°$, and $\widehat{CNC_\alpha} = 123.6°$). Because the shortest distance corresponds to a dihedral angle of $0°$, the PMF at a given force was finally recovered as

$$U_F(\theta) = U_0(\theta) - F \times (L(\theta) - L(0)), \qquad \textbf{[S4]}$$

where θ is the dihedral angle (ϕ or ψ), U_F is the PMF at a force F, U_0 the PMF in the absence of force, and $L(\theta)$ the distance between the first and the last atom for a given value θ of the dihedral angle.

Decaalanine Under Force
As detailed in *Materials and Methods* and mentioned in the text, we have performed an 80-ns unbiased simulation of decaalanine at 30 pN to probe helix stability under force. As shown in Fig. S1, the helix unzips on a nanosecond timescale and its end-to-end distance later fluctuates around high values close to 28 Å. Excursions back to distances corresponding to the folded helix are also observed, suggesting that the corresponding PMF is rather flat under this force. This is confirmed by the direct calculation of the PMF with umbrella-sampling simulations, as shown in the text (Fig. 2D).

Dihedral Angle Dependence of Dipeptide End-to-End Distance
As detailed in *Materials and Methods*, we calculated the distance between the first and the last atom involved in the dihedral angle for each value between $-180°$ and $180°$ to build a simple model to explain the force dependence of dihedral angle distributions. This was done analytically by assuming constant bond lengths and angles that we estimated by averaging over all backbone atoms in the equilibrium simulation of ubiquitin ($d_{C-N} = 1.346$ Å, $d_{N-C_\alpha} = 1.455$ Å, $d_{C_\alpha-C} = 1.522$ Å, $\widehat{NC_\alpha C} = 111.6°$, $\widehat{C_\alpha CN} = 117.6°$, and $\widehat{CNC_\alpha} = 123.6°$). Results are shown in Fig. S2.

Selection of Chemically Unfolded Configurations
Chemically unfolded configuration were selected along a >100-ns trajectory of ubiquitin denaturation in aqueous urea and at high temperature. As detailed in *Materials and Methods*, configurations that were subsequently quenched to 300 K obey two criteria: (*i*) their RMSD is larger than 10 Å with respect to the native state, and (*ii*) they are more than 1 ns apart from each other. In Fig. 3, we show the RMSD (Fig. 3A) and the radius of gyration (Fig. 3B) of the protein along the unfolding trajectory.

Estimation of the Residual Charge at pH = 2
Using the software PROPKA to calculate the individual pK_a of the ubiquitin amino acid side chains (16), we estimate that the protein carries a positive charge of $\sim +13e$ at pH 2, which could lead to mutual repulsion between residues and to a chain extension much less pronounced than that at neutral pH (even if these charges are partly screened by counterions).

Predicting the Experimental Radii of Gyration
Using software such as CRYSOL (17), it is possible to estimate the expected experimental radius of gyration of a protein knowing its atomic coordinates. For the native state of ubiquitin (PDB ID 1UBQ) this leads to a predicted value of 13 Å, whereas the direct measurement in the simulations leads to 11.9 Å. Slightly higher values are also expected for the chemically unfolded configurations.

To avoid any ambiguity we chose to compare the ratios between the folded and unfolded states instead of the absolute values.

Relaxation Time of Chemically Denatured Configurations

As mentioned above, each protein configuration selected along the unfolding trajectory in the presence of urea was later relaxed for 12 ns at ambient temperature. To check that the average backbone properties do not further evolve, we also tried to relax each of these configurations for 12 ns more. As shown in Figs. S4 and S5, the dihedral angle distribution does not change much, suggesting that configurations are equilibrated on this timescale.

1. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–1802.
2. Mackerell AD Jr, Feig M, Brooks CL 3rd (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 25(11):1400–1415.
3. Darden T, Perera L, Li L, Pedersen L (1999) New tricks for modelers from the crystallography toolkit: The particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* 7(3):R55–R60.
4. Lindorff-Larsen K, et al. (2012) Systematic validation of protein force fields against experimental data. *PLoS ONE* 7(2):e32131.
5. Best RB, Buchete NV, Hummer G (2008) Are current molecular dynamics force fields too helical? *Biophys J* 95(1):L07–L09.
6. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14(1):33–38, 27–28.
7. Hua L, Zhou R, Thirumalai D, Berne BJ (2008) Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc Natl Acad Sci USA* 105(44):16928–16933.
8. Wintrode PL, Makhatadze GI, Privalov PL (1994) Thermodynamics of ubiquitin unfolding. *Proteins* 18(3):246–253.
9. Kumar S, et al. (2008) Scalable molecular dynamics with NAMD on the IBM Blue Gene/L system. *IBM J Res Develop* 52:177–188.
10. Bustamante C, Marko JF, Siggia ED, Smith S (1994) Entropic elasticity of lambda-phage DNA. *Science* 265(5178):1599–1600.
11. Frishman D, Argos P (1995) Knowledge-based protein secondary structure assignment. *Proteins* 23(4):566–579.
12. Pace CN, Scholtz JM (1998) A helix propensity scale based on experimental studies of peptides and proteins. *Biophys J* 75(1):422–427.
13. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. *Science* 334(6055):517–520.
14. Darve E, Rodríguez-Gómez D, Pohorille A (2008) Adaptive biasing force method for scalar and vector free energy calculations. *J Chem Phys* 128(14):144120.
15. Darve E, Pohorille A (2001) Calculating free energies using average force. *J Chem Phys* 115:9169–9183.
16. Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pKa values. *Proteins* 61(4):704–721.
17. Svergun DI, Barberato C, Koch MHJ (1995) CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Cryst* 28:768–773.
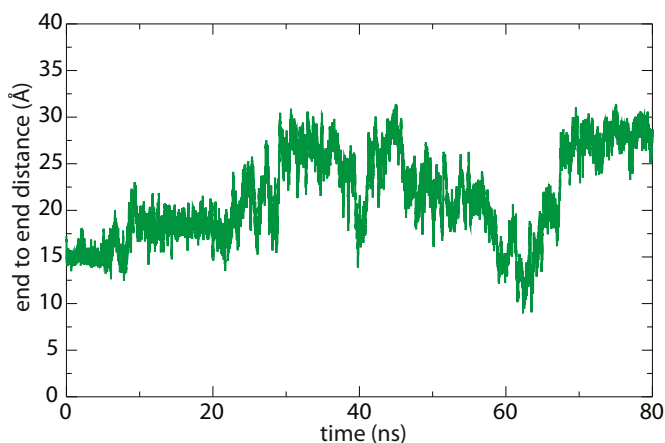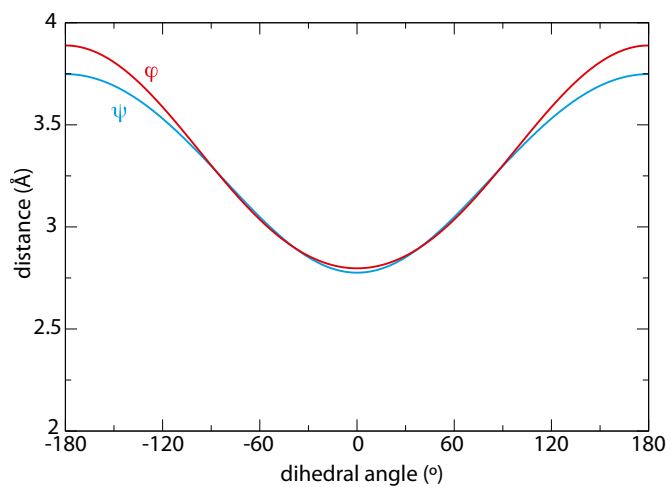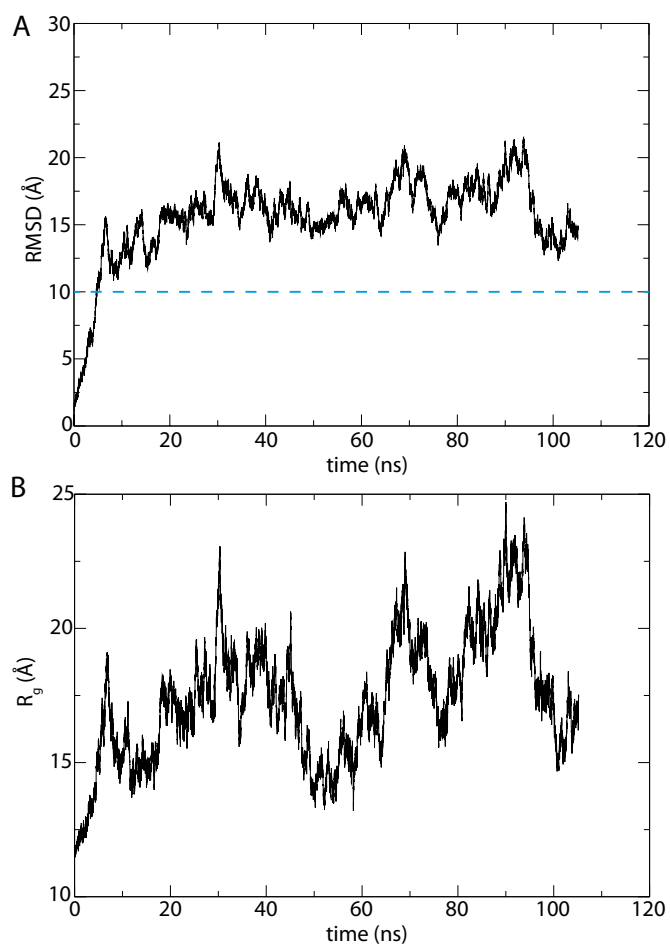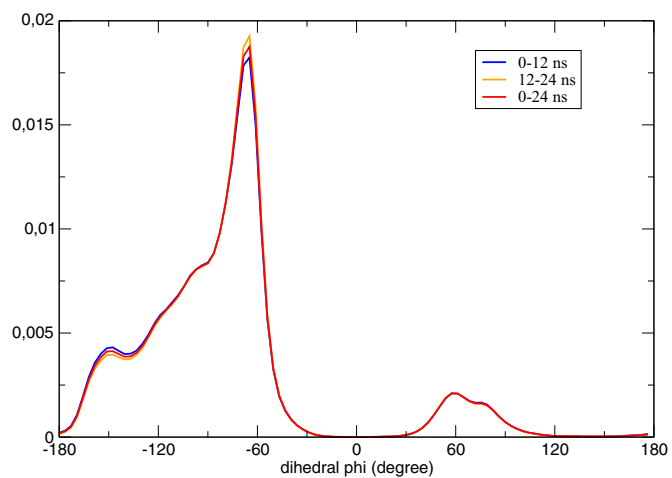
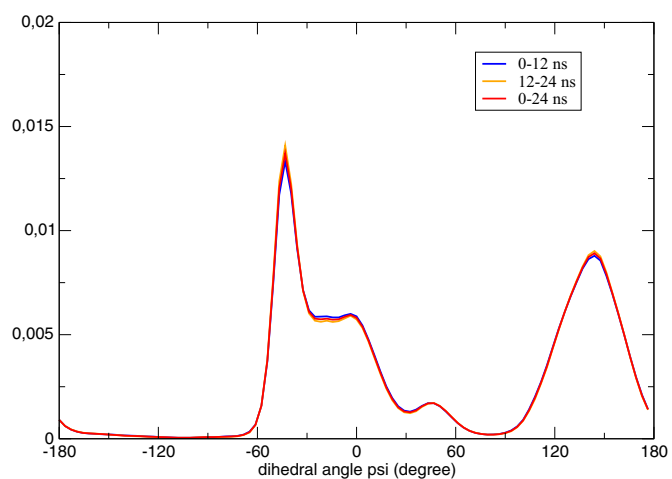**Fig. S1.** Decaalanine under a force of 30 pN. End-to-end distance as a function of time.

Fig. S2. Distance between the first and last atoms involved in a given dihedral angle ϕ or ψ as a function of the dihedral angle value for the dialanine peptide.

Fig. S3. Generation of chemically denatured configurations: RMSD (*A*) and the radius of gyration (*B*) of the protein along the unfolding trajectory. The blue line represents the 10-Å RMSD cutoff that was used.

Fig. S4. Distributions of dihedral angles φ between 0–12, 12–24, and 0–24-ns windows of equilibration for each selected unfolded configuration.

Fig. S5. Distributions of dihedral angles ψ between 0–12, 12–24, and 0–24-ns windows of equilibration for each selected unfolded configuration.