Elasticity, structure, and relaxation of extended proteins under force

Guillaume Stirnemann[a], David Giganti[b], Julio M. Fernandez[b], and B. J. Berne[a,1]

Departments of [a]Chemistry and [b]Biological Sciences, Columbia University, New York, NY 10027

Contributed by B. J. Berne, January 10, 2013 (sent for review November 20, 2012)

Force spectroscopies have emerged as a powerful and unprecedented tool to study and manipulate biomolecules directly at a molecular level. Usually, protein and DNA behavior under force is described within the framework of the worm-like chain (WLC) model for polymer elasticity. Although it has been surprisingly successful for the interpretation of experimental data, especially at high forces, the WLC model lacks structural and dynamical molecular details associated with protein relaxation under force that are key to the understanding of how force affects protein flexibility and reactivity. We use molecular dynamics simulations of ubiquitin to provide a deeper understanding of protein relaxation under force. We find that the WLC model successfully describes the simulations of ubiquitin, especially at higher forces, and we show how protein flexibility and persistence length, probed in the force regime of the experiments, are related to how specific classes of backbone dihedral angles respond to applied force. Although the WLC model is an average, backbone model, we show how the protein side chains affect the persistence length. Finally, we find that the diffusion coefficient of the protein's end-to-end distance is on the order of 10^8 nm^2/s, is position and side-chain dependent, but is independent of the length and independent of the applied force, in contrast with other descriptions.

internal diffusion | potential of mean force | protein elasticity

The development of single-molecule force spectroscopies [atomic force microscopy (AFM), and optical or magnetic tweezers] in the last two decades has opened a whole new and exciting field (1, 2). It is now possible to manipulate biomolecules directly at a molecular level and to study their behavior under force (3). It is therefore not surprising that these techniques have been applied in a broad gamut of contexts, in particular for proteins. In some cases, including enzyme catalysis (4), protein–ligand interaction (5), or folding and unfolding events (6), force is used as a probe, altering the free-energy landscape and the dynamics of the protein (7), and provides valuable kinetic and mechanistic insights. For other systems, force is of direct biological relevance [e.g., cellular adhesion (8) or muscle elasticity (9)].

The elasticity of a polypeptide is typically modeled using the worm-like chain (WLC) (1, 2) model of polymer elasticity. Although this simple model from polymer physics has proven remarkably successful at describing and interpreting experimental data, it lacks molecular details associated with proteins extended under force. However, substrate flexibility to adopt the right geometry is key in many situations, including, e.g., molecular recognition (10) or enzymatic reactivity (11). Therefore, a precise understanding of how force modulates and influences protein flexibility at a single-amino acid level is essential and is not provided by the aforementioned model.

From a dynamical perspective, it is still unclear how applied force affects internal diffusion of a polypeptide along the pulling trajectory. A joint experimental and theoretical study has suggested that increasing force greatly enhances internal friction along the end-to-end coordinate with a power law dependence (12). Indeed force spectroscopy measurements have reported values on the order of 10^2–10^4 nm^2/s (12–14) about five orders of magnitude slower than that obtained for unfolded protein in the absence of force as measured by fluorescence techniques (15, 16). However, we recently suggested that the slow diffusion coefficients of proteins estimated by force spectroscopies are due to the viscous drag on the microscopic objects (beads, tips, surfaces) they are necessarily tethered to (14). Diffusion of untethered protein under force as studied in molecular dynamics (MD) simulations occur on a much faster timescale (typically 10^8 nm^2/s), close to that obtained from fluorescence techniques where the probes are of molecular size. These similarities suggest that force would have a limited impact on diffusion along the longitudinal coordinate. However, the effect of the tether may be of relevance for comparison with diffusion in vivo since force is necessarily exerted with a tethering agent.

Here, we combine MD simulations of a full protein system in explicit solvent to unveil some key molecular aspects of protein relaxation under force. We start by showing that the WLC model successfully explains our simulation data and is also in very good agreement with available experimental results. We then show that the molecular details arising from the dihedral angles lead to WLC behavior. Changes in both the dihedral angles and their fluctuations appear to be necessary and sufficient to give rise to the success of the WLC. Analysis of the molecular dynamics show that the applied force has no influence on the diffusion coefficient along the pulling coordinate (the end-to-end distance) and that the diffusion coefficient of the end-to-end distance is length independent but is position and side-chain dependent.

Results and Discussion

WLC Model and Comparison with Experiments. The WLC model is a very popular and successful model from polymer physics to describe proteins and other biomolecules under force (1, 2, 7). The polymer is considered to be a semiflexible rod whose behavior is uniquely and totally defined by its total length (contour length L_c) and a persistence length p. This latter is a measure of the polymer local stiffness, i.e., the length scale above which the rod can be curved. Proteins are relatively flexible polymers whose persistence length is usually on the order of a nanometer (6, 17), which corresponds to the length scale of a single peptide unit.

We first verify that our simulations agree with prediction from the WLC model. Trajectories of all-atom ubiquitin in explicit solvent at 300 K are generated at different forces as described in the *Materials and Methods*. Examples of such trajectories are given in Fig. 1*A*, which shows both equilibrium fluctuations of the end-to-end distance at 250 pN and how the protein shrinks when the force is reduced (force-quenched) to 100 pN. At each force, including others not depicted in Fig. 1*A*, equilibrium properties are averaged over tens of nanoseconds once plateauing of the end-to-end distance is observed. Typically, this takes between 3 (at 100 pN) and 60 ns (at 30 pN) after the force has been quenched. Snapshots of the protein at representative forces are shown in Fig. 1*B*, suggesting that, although

Author contributions: G.S., D.G., J.M.F., and B.J.B. designed research; G.S. performed research; G.S., D.G., J.M.F., and B.J.B. analyzed data; and G.S. and B.J.B. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: bb8@columbia.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1300596110/-/DCSupplemental.

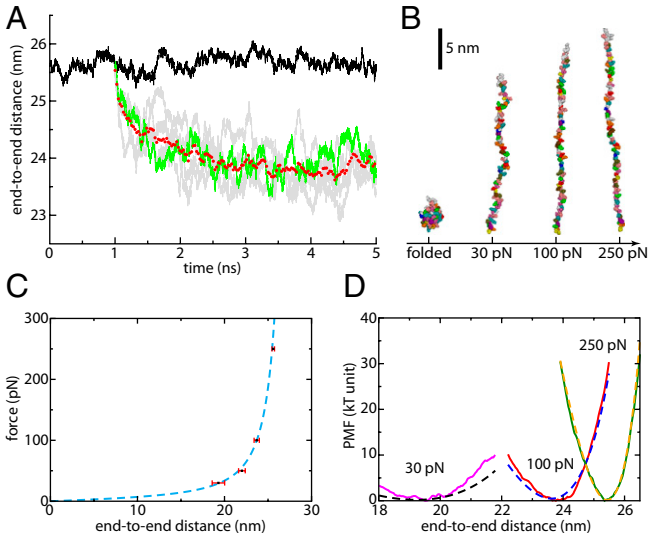

Fig. 1. (*A*) End-to-end distance as a function of time for ubiquitin under an applied force of 250 pN (black curve). Relaxation trajectories are obtained by quenching force from 250 to 100 pN (gray curves, one highlighted in green) and averaged over five such trajectories (red circles). Equilibrium data are then accumulated once the average end-to-end distance has reached a plateau. A similar procedure is applied at other forces. (*B*) Snapshots of ubiquitin in a folded configuration (*Left*) and at different forces (30, 100, and 250 pN). Solvent molecules are not represented. (*C*) Force-extension profile at each force (black squares, average; red bars, standard deviation) and the corresponding WLC fit (dashed blue line). (*D*) PMF as a function of end-to-end distance at different forces (30 pN, magenta curve; 100 pN, red curve; 250 pN, green curve) and comparison with the WLC predictions (30 pN, dashed black curve; 100 pN, dashed blue curve; 250 pN, dashed yellow curve).

no secondary structure is formed, the protein is not yet totally extended and exhibits some lateral fluctuations and local structure.

We show in Fig. 1*C* that the average end-to-end distances L obtained at different forces down to 30 pN, are well fitted by a WLC chain force-extension profile (see expression in *Materials and Methods*) with $L_c = 28.4$ nm [corresponding to a peptide unit's length of 0.38 nm, in agreement with the experimental estimate of 0.4 ± 0.02 nm (18)] and $p = 0.39$ nm. Because the fit of the force-extension data can lead to uncertainties in the determination of the total contour length of the protein, we also use an alternate technique to estimate the peptide unit's length that leads to the same value of 0.38 nm (*SI Text* and Fig. S1). The persistence length value of 0.39 nm for unfolded ubiquitin is in remarkable agreement with that obtained by AFM experiments, including some on the same protein (2, 19, 20), which suggest that simulations reach equilibrium. However, the WLC model alone is expected to break down at forces lower than 10–20 pN for then the barrier to collapse to a more compact and stable state should be low enough that hopping will be observed in the experiment on the timescale of seconds (21). For such low forces, protein simulations on a nanosecond or even a microsecond timescale might see the protein get trapped in a metastable state. As an illustration, we repeated our simulations applying a force of 15 pN and did not observe any convergence of the end-to-end distance even after 150 ns of simulation.

As force decreases, fluctuations of the end-to-end distance around its average value become more important (Fig. 1*C* and Fig. S2). In Fig. 1*D* we further compare the potential of mean force (PMF) as a function of the end-to-end distance determined in our simulations (using the umbrella sampling technique) with the WLC predictions using the parameters determined independently from the force-extension profile of Fig. 1*C*. A very good agreement is found at 250 pN but the PMF increasingly deviates from the WLC prediction when force is decreased. We also note that no roughness is observed in the free-energy profile (PMF) at high applied force—the PMF is very smooth. This observation contrasts with a recent suggestion that local roughness of at least $4k_BT$ could offer an explanation to the very slow effective diffusion coefficient of the end-to-end distance in this regime of forces and extensions (12). At worst, roughness on the order of $1k_BT$ starts appearing at the lowest force investigated, which may be explained by the existence of contacts between adjacent side chains, a feature absent from the simple WLC model (7).

The overall similarities between our simulations and the WLC model, both in the force-extension profile and in the PMF along the end-to-end distance, clearly demonstrate the applicability of the model, especially at high stretches. The absence of significant roughness suggests that the projection onto the end-to-end coordinate provides a dynamically relevant energy landscape.

Finally, one could argue that, because of the limited timescale readily accessible to our simulations, the protein could be "trapped" in a metastable state and that the end-to-end distance could further decay toward its true average value if much longer simulations were performed. Three complementary arguments suggest convergence is reached: our force-extension results agree very well with experimental data, the average end-to-end distance corresponds to the true minimum of the PMF at each force (assuming there is no trapping in the sampling of the PMF), and the timescale of internal diffusion (as discussed later) is fast enough so that a good sampling is ensured on a nanosecond timescale.

Structural Aspects of Protein Relaxation. In the simulations of ubiquitin acted upon by forces ranging from 30 to 250 pN, no secondary structure and no native contacts between residues were observed. This is consistent with the sudden increase in length observed experimentally upon unfolding of a protein under mechanical force, which generally corresponds to the release of the totality of its amino acids (19): proteins are thus almost fully extended in this force regime. It is, however, evident from Fig. 1 that proteins at different forces exhibit distinct structural features that must be implicated in the variations of the end-to-end distance under force.

We thus examine the effect of force on the backbone degrees of freedom of a protein. Bond lengths and bend angles appear to be insensitive to the subnanonewton forces usually used in experiments (*SI Text* and Fig. S3). This observation may not be surprising given the magnitude and the stiffness of the associated potentials. Hence a force of 1 nN is typically required to stretch a bond (or a bend angle) by $\lesssim 5\%$. (Investigation of such high-energy effects would require more accurate force fields for bond bending and bond stretching, but such effects are not our main focus here.) For comparison, ubiquitin at 800 pN is more than 35% longer than it is at 30 pN. Thus, the elastic response probed in most force spectroscopy experiments does not involve the elongation of bonds and bend angles.

In contrast, the backbone dihedral angles usually defined as $\phi = C_{(-1)}NC_\alpha C$ and $\psi = NC_\alpha CN_{(+1)}$ are dramatically affected by force (the third backbone dihedral angle $\omega = C_{\alpha(-1)}C_{(-1)}NC_\alpha$ is readily close to 180° in the absence of force because of the planarity imposed by conjugation). The Ramachandran plots presented in Fig. 2*A* illustrate these large differences. 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 Protein Data Bank (PDB) (22) grow, mainly in the polyproline II region approximately (−75°, 150°), the region that is also populated in chemically unfolded proteins (23). Nevertheless, we again stress that no secondary structure is formed.

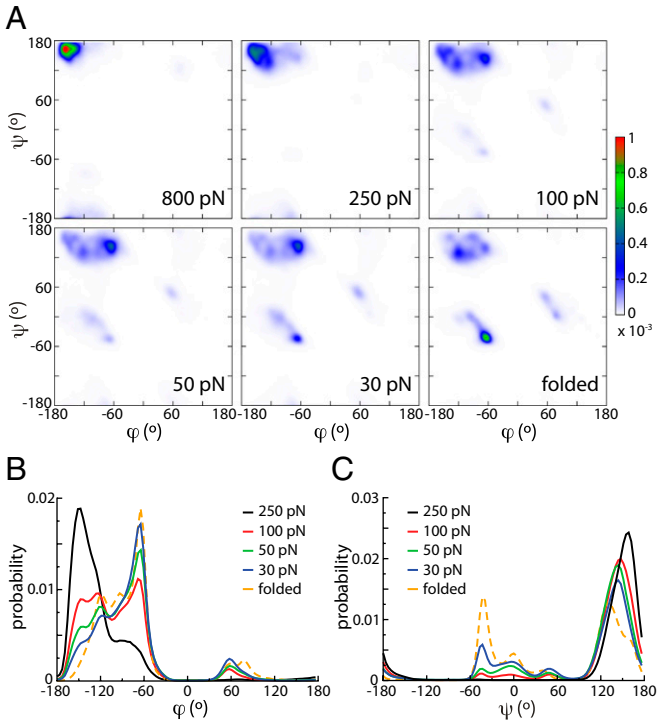

Fig. 2. (*A*) Ramachandran plots of ubiquitin (*Upper*, from *Left* to *Right*: 800 pN, 250 pN, 100 N; *Lower*, from *Left* to *Right*: 50 pN, 30 pN, and folded at no force). (*B*) Distribution of the ϕ dihedral angle for unfolded ubiquitin at different forces and compared with the distribution for the folded, equilibrium protein at no force. (*C*) Same plot for the ψ dihedral angle.

Changes are more prominent if we consider the distributions of angles ϕ (Fig. 2*B*) and ψ (Fig. 2*C*), corresponding to projections of the Ramachandran plot onto the respective axes. The two angles behave differently with force. ϕ is definitely the more sensitive (Fig. 2*B*): when force is lowered, the peak close to −180° progressively shifts to higher values while disappearing, and sharp peaks around −60° and 60° (similar in terms of the corresponding peptide end-to-end distance) that are absent at 250 pN rise. It is remarkable that the distribution at the lowest force investigated here (30 pN) is very similar to that of the folded structure in the absence of force. However, although the distributions are similar, they are very different in nature, as shown in *SI Text*. For the folded state, the heterogeneous distribution arises from static heterogeneities among the amino acid sequence, whereas under force, most of the amino acids along the sequence sample significant part of the distribution, which is therefore due to dynamic disorder (Fig. S4) (the same observation is made for the ψ angle). The force dependence of ψ is more subtle (Fig. 2*C*) and differs from that of ϕ in some key aspects. First, a sharp peak is present close to 180° at any force although its position shifts and its amplitude decreases as force is decreased. Even if smaller angles are found at the smaller forces, the distribution at 30 pN is markedly different from that of the folded protein. In particular, the sharp peaks between −60° and 0°, leading to shorter peptide end-to-end distance is not totally present (Fig. 2*A*).

Although major variations of dihedral angles are observed upon relaxation of the force, it is still unclear at this point whether this is a cause or a consequence of the relaxation of the end-to-end distance. We now investigate whether artificially altering the dihedral free-energy surfaces in the MD simulations leads to a different WLC behavior. In realistic force fields, different types of interactions contribute to the dependence of the free-energy profile on the dihedral angles. One of them accounts for the regular nonbonded (van der Waals and electrostatic) interactions between the different atoms involved. However, this alone usually fails to reproduce the specific dihedral angle potential obtained from more precise, ab initio calculations. Some cosine-based corrections are therefore added to give better agreement with high level theory. In most cases, the magnitude of these corrections is small (~0–1 kcal/mol) with respect to the contribution of electrostatic and van der Waals interactions. To explore the specific effect of ϕ and ψ backbone dihedral angles on the WLC behavior, we artificially modify the corresponding dihedral potentials in our simulations. To introduce a significant perturbation, we have therefore repeated our simulations using either a ψ or a ϕ potential 10 times larger than the unperturbed one (*SI Text*). Although of similar magnitude, these potentials fundamentally differ in that sense that they bias the angle toward 0° for ϕ but toward ±180° for ψ. The resulting end-to-end distances at two representative forces and the corresponding WLC fits are shown in Fig. 3. For this fit, we fix the contour length to be the same than that determined for the unperturbed system (28.4 nm), the persistence length being the only free parameter. To reduce the computational effort, we only repeated our simulations at 100 and 250 pN so that the results of these fits should be considered as qualitative. For high ϕ dihedral potentials, we find a persistence length $p = 0.19$ nm, which is one-half that of the unperturbed protein ($p = 0.39$ nm): in this case, the protein is more flexible because of the bias toward $\phi = 0°$. On the contrary, the persistence length if using high ψ dihedral potentials is $p = 0.59$ nm, showing that the polypeptide is stiffer in this case. These results unambiguously show that the WLC behavior and thus protein flexibility are intimately tied to the free-energy profile along the ψ and ϕ dihedral angles. Artificial alteration of the corresponding potentials leads to different values of the persistence length.

The above study demonstrates how the applied force affects the dihedral angle distributions and through this the protein flexibility. The protein collapse upon force-quench is found to be due to a decrease in the average backbone dihedral angles whose trajectories in the Ramachandran space correlate well with that of the end-to-end distance (Fig. S5). These dihedral angles are also responsible for the fluctuations of the end-to-end length at a given force (Figs. S2 and S5).

Additional confirmation of the key role played by dihedral angles comes from the effect of side chains on protein flexibility. Earlier studies showed that a purely entropic chain analog of ubiquitin, where all attractive interaction between the protein atoms was removed, is much stiffer, with a persistence length of 1.2 nm (24) (three times larger than that of regular ubiquitin). Indeed the distribution of a given dihedral angle mainly result from both energetic and steric interactions between the atoms or groups of atoms defining this angle. We have examined the variations of ϕ and ψ for each residue independently (*SI Text* and Fig. S6). In

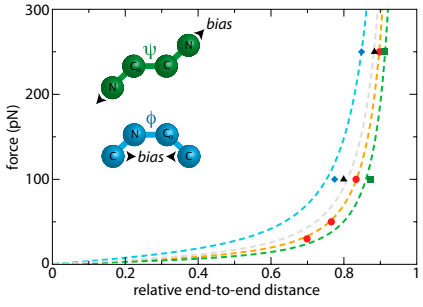

Fig. 3. Effect of dihedral angles and side chains on the WLC behavior (red, regular potentials; blue, biasing the ϕ potential toward 0°; green, biasing the ψ dihedral potential toward 180°; black, polyglycine analog of ubiquitin); points represent simulation data, and dashed lines show the best WLC fits using $L_c = 28.4$ nm.

most cases, fluctuations are broad and poorly correlated with the size of the side chains. Only two amino acids are found to behave very differently: glycine and proline. The proline side chain is involved in the backbone through a five-atom cycle, and it is therefore well known that it constrains dihedral orientations. For example, its ϕ remains always close to −60° for any force. For glycine, major differences are observed and both angles can explore lower angular values more easily because of the absence of a side chain. As a consequence, the end-to-end C–C distance for each glycine is more sensitive to force, and thus more flexible. Similar observations were made for short protein fragments whose flexibility is sensitive to the glycine content (25). To confirm this particular behavior, we performed simulations of an all-glycine analog of ubiquitin, i.e., a protein containing the same number of residues but all mutated to glycine. The resulting end-to-end distances as a function of applied force are shown in Fig. 3. As predicted, this homopolymer is more flexible with $p = 0.31$ nm, which is ∼25% shorter than that of the wild-type ubiquitin. At any force, the Ramachandran plots of polyglycine are also very different from that of ubiquitin (Fig. S7). Thus, a direct connection is made between this increased flexibility and the different distributions of dihedral angles, again showing the strong interplay between these two aspects as well as the role of side-chain interactions (both energetic and steric) in determining the persistence length.

Here, we show how dihedral angles determine a persistence length $p \sim 0.4$ nm for an unfolded protein. This observation may seem in contrast with some force-ramp and constant-velocity measurements showing a static disorder of p on a second timescale among different collapse or extension trajectories (17, 26), especially at the lower forces. Possible explanations include the effect of the ionic environment of the experimental buffer solutions that could alter the dihedral free-energy profile, or the strong effect of tethering on the protein internal diffusion (14) that could result in different pathways in the Ramachandran space and in out-of-equilibrium states.

Dynamical Aspects of Protein Relaxation. We now investigate how internal diffusion (i.e., diffusion along the end-to-end distance) in ubiquitin is affected by applied force and the resulting extension. The fast, nanosecond relaxation observed upon force-quench (Fig. 1*A*) at any force suggests that diffusion occurs on a very fast $D \sim 2 \times 10^8$ nm^2/s timescale (using a very approximate estimation detailed in *SI Text*, later confirmed by more precise calculations as detailed below). These values are in close agreement to that measured in fluorescence techniques (15, 16), but five to six orders of magnitude faster than observed in atomic force spectroscopy measurements (12–14). We and collaborators have recently suggested that the observed slow diffusion coefficients seen in force spectroscopies experiments result from the viscous drag on the microscopic objects (beads, tips, surfaces) they are necessarily tethered to (14), and do not correspond to an intrinsic property of the protein. Diffusion is much faster in the simulations because in the usual steered MD setup force is directly applied to one extremity of the protein, the other one being fixed, without any intermediate tethering agent. This conclusion contrasts with that of a previous attempt to separate the dynamics of a polypeptide from that of the AFM tip it is tethered to where measured diffusion coefficients were found to be as much as six orders of magnitude smaller than those measured in bulk solution without tethers (12). These values were interpreted in terms of a frictional WLC model, predicting a power-law dependence of friction on force as $F^{3/2}$, and also that friction is inversely proportional to the protein contour length L_c. Fair agreement was claimed to be found between this model and the experimental data (12). These results are surprising because one would expect that force would only affect the free-energy surface on which the protein is moving, not the diffusion dynamics itself. Interestingly, another popular model [the Rouse model (27)] has been recently applied to interpret simulation results on diffusion dynamics of short polypeptides under force (28) and that of unfolded proteins in Förster resonant energy transfer experiments (29). The protein is modeled as a chain of N beads connected by springs. Although such aspect was not the focus of these previous works, this model predicts that diffusion along the end-to-end distance slows down as N increases, i.e., as L_c increases.

Here, we use a method described previously (30, 31) to estimate D at different positions and forces (*SI Text* and Fig. S8). Briefly, an additional harmonic potential bias is used to constrain the end-to-end distance L around a target value L_T. Under the assumption that the free-energy surface is locally harmonic, which is ensured by using a bias potential that is stiff enough with respect to the actual PMF, $D(F, L_T)$ is recovered from the autocorrelation function $\delta L = L - \langle L \rangle$ as $D = \langle \delta L^2 \rangle^2 / \int_0^\infty \langle \delta L(0) \delta L(t) \rangle dt$.

The evolution of $D(F, L_{eq}(F))$ as a function of force, evaluated at the equilibrium position at each force $L_{eq}(F)$ (Fig. 1*B*), is shown in Fig. 4*A*. We note a very weak dependence of D upon force: when force is multiplied by ∼10, from 30 to 250 pN, D is only reduced by less than 30%, in strong contrast with the model and the measurements mentioned above (12), which would predict a difference by a factor ∼25. In addition, these surprisingly low experimental values of the internal diffusion coefficient were attributed to the roughness of the free-energy surface (12). As predicted by Zwanzig (32), local barriers of $\sim 4 k_B T$ would be required to explain these low effective diffusion coefficients (12). Our results rather suggest that the PMF is smooth and that no large local barriers are present, especially at high forces (Fig. 1*C*). As a consequence, surface roughness can probably be ruled out as an explanation of these large differences. The low values of experimentally measured diffusion coefficients may be due to the effect of the tip, which was not totally removed in the analysis of the data and has been shown to lead to such low values of the diffusion coefficient (14).

Instead, we propose that the observed (weak) force dependence at the respective equilibrium end-to-end distances is in fact a position dependence. When we repeat the measurement for $L = 25.5$

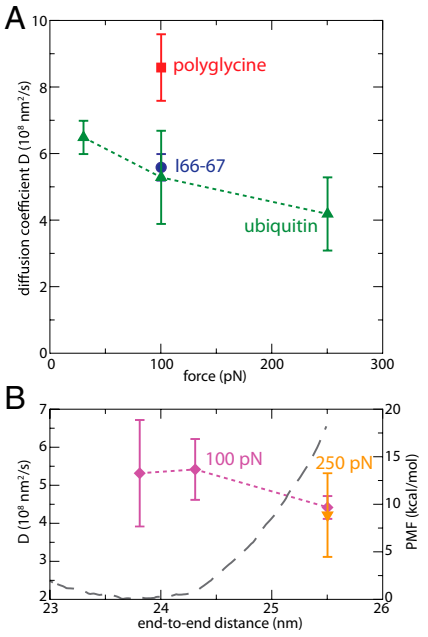

Fig. 4. Diffusion coefficients D along the end-to-end distance. (*A*) D as a function of force, for regular ubiquitin (green triangles), its polyglycine analog at 100 pN (red square) and the longer I66-67 (blue circle) at 100 pN. Here, D is calculated for each force at the corresponding average end-to-end distance L. Error bars are estimated after block averaging. (*B*) D as a function of length at 100 pN (violet diamonds); D at $F = 250$ pN and $L = 25.5$ nm, the equilibrium position at this force, is given for comparison. The PMF at 100 pN is also shown in gray dashed line.

nm (equilibrium distance at 250 pN) and $F = 100$ pN (Fig. 4*B*), the respective diffusion coefficients at 100 and 250 pN are very similar, suggesting that, for a given protein configuration, increasing force does not lead to a larger friction coefficient. Force's effect on friction is rather indirect, simply altering the PMF (Fig. 1*D*) and subsequently the average end-to-end distance. In the absence of force, the extent of the variations of D along the PMF describing folding and unfolding is still controversial, although it was shown to strongly depend on the coordinate that was considered (33). However, the observed changes are often smaller than one order of magnitude. Some studies have suggested that these changes occur because of the formation of native contacts such as hydrogen bonds, locally increasing the friction around specific positions (28, 33). For an extended polypeptide under force, we do not observe any strong position dependence over the limited range accessed by the protein (Fig. 4*A*): L varies between ∼20 and ∼25 nm over the 30- to 250-pN force range. At a given force, a similar weak dependence is observed along the PMF (Fig. 4*B*), mostly lying within error bars. These results are in agreement with our prior observation that neither secondary structure nor native contacts are formed at these forces.

To investigate the length dependence of the diffusion coefficient, we have performed simulations of a 188-residue fragment of the giant protein titin consisting of the two domains I66 and I67 (9) using the same setup as already discussed for ubiquitin. Ig domains of titin have been extensively studied by force spectroscopy techniques (2, 9) and we have chosen this system both for convenience and to show that our conclusions hold for a protein very different from ubiquitin. In particular, this unfolded fragment is much longer, reaching an average end-to-end distance $L_{eq}(250\text{ pN}) = 63.0$ nm at 250 pN. The resulting value of D is shown in Fig. 4*A* and is seen to be very similar to that of ubiquitin at the same force: internal diffusion does not appear slowed down as the polymer length is increased, in strong contrast with earlier predictions (12).

The diffusion coefficients ($\sim 5 \times 10^8$ nm^2/s) reported here are consistent with the value of $\sim 8 \times 10^8$ nm^2/s for the end-to-end diffusion of a short polypeptide in the absence of force and described by the same force field (34). Once corrected for threefold lower viscosity of our water model, our diffusion coefficients also closely agree with the estimate of 1.8×10^8 nm^2/s from triplet-state contact quenching (16). Finally, the values obtained using very stiff biasing potentials are in fair agreement with the approximate estimation from the collapse trajectories of $\sim 2 \times 10^8$ nm^2/s (*SI Text*). Now how can we relate the measured value for D with the importance of dihedral angles and side chains in providing flexibility to the protein? Because relaxation of an extended polypeptide upon force-quench is largely due to large changes in the dihedral angles, as shown above, diffusion along the end-to-end distance is likely to correspond to diffusion along the dihedral coordinates. Indeed, our results are in surprisingly good agreement with previous calculations on the dialanine dimer in water (30), where diffusion along the ω ($C_{\alpha(-1)}C_{(-1)}NC_{\alpha}$) dihedral angle was found to be almost position independent and close to ∼0.2 rad^2/ps. Once converted in terms of end-to-end distance using the relationship $D(L) = D(\omega) \times (dL/d\omega)^2$, the values correspond to $\sim 1\text{–}5 \times 10^8$ nm^2/s in the 180–90° range, reaching a maximum at 90°. Although the system is different and another dihedral coordinate has been considered, the agreement is remarkable. In particular, the small increase in D when L decreases coincides with the increase in $(dL/d\omega)$ in the 180–90° range for a fixed value of diffusion along the dihedral coordinate. Therefore, the internal diffusion of protein under force shows striking similarities with the diffusion along dihedral coordinates. A direct consequence is that the importance of side chains is nonnegligible: diffusion of the polyglycine analog of ubiquitin is almost twice as fast as that of regular ubiquitin at similar extension and force (Fig. 4*A*).

Concluding Remarks

In this paper, we combine molecular dynamics simulations of a full protein system in explicit solvent to uncover some key molecular aspects of protein elasticity under force. We first show that, at forces $F \geq 30$ pN, the WLC model successfully explains the simulation data, especially at high applied force. Both the contour length and the persistence length are in very good agreement with available experimental data. We find evidence for a very strong interplay between the protein elasticity and the response of the backbone dihedral angles to force, and show how this interplay leads to the success of the WLC model. We provide a sensitivity analysis of different classes of dihedral angles to force, and show how side chains affect the persistence length of proteins through their effect on the exploration of backbone dihedral space. For example, we show that a polyglycine analog is ∼25% more flexible than a regular protein. Finally, artificial alteration of the dihedral potentials in the simulations is found to significantly affect the persistence length.

From a dynamical perspective, we show that the applied force should have practically no influence on the diffusion coefficient along the pulling coordinate (the end-to-end distance) in force spectroscopy experiments. We show that this internal diffusion coefficient is position and side-chain dependent, but length independent. Thus, applied force does not dramatically affect intramolecular friction, but does change the free-energy surface (PMF) the protein is moving along. We also show that diffusion is several orders of magnitude faster for untethered proteins in simulations than for tethered proteins (as studied experimentally), as we suggested earlier (14).

Since being introduced to treat experimental data from the first force spectroscopy measurements (1, 2), the WLC model has been very successful in a wide variety of systems and contexts. This study provides an explanation of why the WLC model has enjoyed such success based on molecular details. Our detailed analysis shows how the dihedral angles determine the protein's persistence length and diffusion dynamics along the end-to-end distance. Protein elasticity and dynamics, especially the timescale of fluctuations along the end-to-end distance, depends crucially on how protein dihedral angles explore key regions of the Ramachandran space. Because the length scale of dihedral motions is that of a single peptide unit, it is therefore not surprising that proteins are very flexible polymers, in contrast with, e.g., nucleic acids.

Because force can modulate flexibility, this molecular picture can help complement and provide an understanding of the experimental data that simple continuous models fail to explain. Several studies have shown that when force is applied to the proteins in certain enzyme–substrate complexes, the relaxation is very sensitive to strength of the applied force, sometimes in a nonlinear fashion (4). An analysis of the kind presented here might shed light on this sensitivity.

Materials and Methods

Molecular Dynamics. All-atom steered MD simulations of wild-type ubiquitin (PDB ID code 1UBQ) and I66-67 protein fragment (PDB ID code 3B43) in explicit solvent were carried out with the software NAMD 2.8, using the CHARMM27 force field for the protein and the TIP3P-CHARMM water model. Details about the system preparation, equilibration, calculation of the PMF, and diffusion coefficients are given in *SI Text*.

WLC Model. A WLC is fully described 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 Eq. **1** as follows:

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

and the PMF force along the end-to-end distance $W(L)$ by the following:

$$W(L) = \frac{k_B T}{p}\left(\frac{L_c}{4}\left[\frac{1}{1-\frac{L}{L_c}} - 1\right] - \frac{L}{4} + \frac{L^2}{2L_c}\right) - F \times L. \qquad [2]$$

ACKNOWLEDGMENTS. We thank Ronen Berkovich, Glen Hocky, and Joe Morrone (Columbia University) for fruitful discussions and comments on the manuscript. This work was supported by grants from the National Institutes of Health [NIH-GM4330 (to B.J.B.)] and in part by the National Science Foundation through TeraGrid resources provided by National Center for Supercomputing Applications (MCA08X002).

1. Bustamante C, Marko JF, Siggia ED, Smith S (1994) Entropic elasticity of lambda-phage DNA. *Science* 265(5178):1599–1600.
2. 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.
3. Neuman KC, Nagy A (2008) Single-molecule force spectroscopy: Optical tweezers, magnetic tweezers and atomic force microscopy. *Nat Methods* 5(6):491–505.
4. Perez-Jimenez R, et al. (2009) Diversity of chemical mechanisms in thioredoxin catalysis revealed by single-molecule force spectroscopy. *Nat Struct Mol Biol* 16(8): 890–896.
5. Merkel R, Nassoy P, Leung A, Ritchie K, Evans E (1999) Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature* 397(6714):50–53.
6. Schlierf M, Li H, Fernandez JM (2004) The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc Natl Acad Sci USA* 101(19): 7299–7304.
7. Berkovich R, Garcia-Manyes S, Urbakh M, Klafter J, Fernandez JM (2010) Collapse dynamics of single proteins extended by force. *Biophys J* 98(11):2692–2701.
8. Thomas W (2008) Catch bonds in adhesion. *Annu Rev Biomed Eng* 10:39–57.
9. Hsin J, Strümpfer J, Lee EH, Schulten K (2011) Molecular origin of the hierarchical elasticity of titin: Simulation, experiment, and theory. *Annu Rev Biophys* 40:187–203.
10. Rini JM, Schulze-Gahmen U, Wilson IA (1992) Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science* 255(5047):959–965.
11. Benkovic SJ, Hammes-Schiffer S (2003) A perspective on enzyme catalysis. *Science* 301(5637):1196–1202.
12. Khatri BS, et al. (2008) Internal friction of single polypeptide chains at high stretch. *Faraday Discuss* 139:35–51, discussion 105–128, 419–420.
13. Kawakami M, Byrne K, Brockwell DJ, Radford SE, Smith DA (2006) Viscoelastic study of the mechanical unfolding of a protein by AFM. *Biophys J* 91(2):L16–L18.
14. Berkovich R, et al. (2012) Rate limit of protein elastic response is tether dependent. *Proc Natl Acad Sci USA* 109(36):14416–14421.
15. Haas E, Katchalskikatzir E, Steinberg I (1978) Brownian-motion of ends of oligopeptide chains in solution as estimated by energy-transfer between chain ends. *Biopolymers* 17(1):11–31.
16. Lapidus LJ, Steinbach PJ, Eaton WA, Szabo A, Hofrichter J (2002) Effects of chain stiffness on the dynamics of loop formation in polypeptides. appendix: Testing a 1-dimensional diffusion model for peptide dynamics. *J Phys Chem B* 106:11628–11640.
17. Walther KA, et al. (2007) Signatures of hydrophobic collapse in extended proteins captured with force spectroscopy. *Proc Natl Acad Sci USA* 104(19):7916–7921.
18. Ainavarapu SRK, et al. (2007) Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophys J* 92(1):225–233.
19. 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.
20. Carrion-Vazquez M, et al. (2003) The mechanical stability of ubiquitin is linkage dependent. *Nat Struct Biol* 10(9):738–743.
21. 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.
22. 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.
23. Jha AK, Colubri A, Freed KF, Sosnick TR (2005) Statistical coil model of the unfolded state: Resolving the reconciliation problem. *Proc Natl Acad Sci USA* 102(37): 13099–13104.
24. Gräter F, Heider P, Zangi R, Berne BJ (2008) Dissecting entropic coiling and poor solvent effects in protein collapse. *J Am Chem Soc* 130(35):11578–11579.
25. Cheng S, Cetinkaya M, Gräter F (2010) How sequence determines elasticity of disordered proteins. *Biophys J* 99(12):3863–3869.
26. Li H, et al. (2001) Multiple conformations of PEVK proteins detected by single-molecule techniques. *Proc Natl Acad Sci USA* 98(19):10682–10686.
27. Rouse PE (1953) A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. *J Chem Phys* 21:1272–1280.
28. Schulz JCF, Schmidt L, Best RB, Dzubiella J, Netz RR (2012) Peptide chain dynamics in light and heavy water: Zooming in on internal friction. *J Am Chem Soc* 134(14): 6273–6279.
29. Soranno A, et al. (2012) Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proc Natl Acad Sci USA* 109 (44):17800–17806.
30. Hummer G (2005) Position-dependent diffusion coefficients and free energies from bayesian analysis of equilibrium and replica molecular dynamics simulations. *New J Phys* 7:34.
31. Straub JE, Borkovec M, Berne BJ (1987) Calculation of dynamic friction on intramolecular degrees of freedom. *J Phys Chem* 91:4995–4998.
32. Zwanzig R (1988) Diffusion in a rough potential. *Proc Natl Acad Sci USA* 85(7): 2029–2030.
33. Best RB, Hummer G (2010) Coordinate-dependent diffusion in protein folding. *Proc Natl Acad Sci USA* 107(3):1088–1093.
34. Yeh IC, Hummer G (2002) Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent. *J Am Chem Soc* 124(23):6563–6568.

Supporting Information

Stirnemann et al. 10.1073/pnas.1300596110

SI Text

Molecular Dynamics Simulations. *Generalities.* All-atom simulations in explicit solvent were carried out with the software NAMD 2.8 (1), using the CHARMM22 force field with CMAP corrections for the protein (2), and the TIP3P-CHARMM water model. We used periodic boundaries conditions and a cutoff of 12 Å for electrostatic and Lennard–Jones interactions. Long-range electrostatic interactions were calculated using the particle mesh Ewald 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. Steered MD simulations of wild-type ubiquitin [Protein Data Bank (PDB) ID code 1UBQ] 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. A similar setup as that described below was used for simulations of the larger I66-67 protein (PDB ID code 3B43) or that of the polyglycine analog of ubiquitin. Overall, the trajectories used in this work represent a total simulation time of ∼0.8 μs.

System preparation. To unfold the protein, we first pull 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 water box module of VMD (4) 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.

Equilibration. The protein was then equilibrated for 6 ns at 250 pN in the isothermal-isobaric ensemble (NPT) 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 Nose–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 more nanoseconds 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 100 ns at this force). At each force, the average end-to-end distance no longer evolved during these production runs, yet fluctuations were observed (Fig. S2). All these trajectories were used for subsequent structural analysis. No dynamical data were extracted from these simulations because of possible bias introduced by the temperature and pressure control.

Collapse simulations. Initial configurations for collapse from 250 to 100 pN were chosen along the 25-ns trajectory at 250 pN. They were propagated in the microcanonical ensemble for 5 ns to avoid spurious effects from pressure and temperature control on the dynamics of collapse. A time step of 1 fs was used. We performed five such simulations to obtain the average relaxation. Because of the large system size, average temperature and pressure along these trajectories are very close to that targeted during the NPT equilibration.

Potential of mean force calculations. At each force, we used umbrella sampling along the end-to-end distance L to recover the corresponding potential of mean force (PMF). Two different values for the sampling frequency along L and the force constant of the constraining potential were used depending on the local PMF stiffness: whereas we sampled every 1 Å with a force constant of 10 kcal·mol^{-1}·Å^{-2} at the higher forces, a sampling every 2 Å with a force constant of 2.5 kcal·mol^{-1}·Å^{-2} was used at lower forces. In both cases, we obtained substantial overlap between adjacent windows. At each L, an equilibration simulation was first performed for 1 ns before a production run was propagated for 2 ns, and used for subsequent analysis. The PMF was finally reconstructed using the program WHAM (http://membrane.urmc.rochester.edu/content/wham).

Modified dihedral potentials. In realistic force fields (CHARMM, OPLS, AMBER, etc.) like the one used here, two different types of interactions contribute to the dependence of the free-energy profile on the dihedral angles connecting atoms numbered 1–2–3–4. One of them accounts for the regular nonbonded (van der Waals and electrostatic) interactions between atoms 1 and 4 and neighboring groups (side chains, etc.). However, this alone usually fails to reproduce the specific dihedral angle potential obtained from more precise, ab initio calculations. Some cosine-based corrections are therefore added to give better agreement with high-level theory. In general, a potential,

$$V_{\mathrm{dih}}(\theta) = K_{\mathrm{dih}}(1+\cos(n\theta-\theta_0)), \qquad \text{[S1]}$$

is used, where K_{dih} is the amplitude of the correction, n is the multiplicity, and θ_0 is a phase term. In the force field that we used (CHARMM 22 with CMAP corrections), there is an additional term for each peptide $V_{\mathrm{CMAP}}(\phi, \psi)$ describing cross-correlations between dihedral angles ψ and ϕ (2). To illustrate the importance of dihedral angles in the chain stiffness and relaxation, we have performed simulations using K_{dih} values 10 times larger than the unperturbed one. All dihedral potentials for the corresponding backbone dihedral angle (either ψ or ϕ), which depend on the nature of the residue, were modified. However, the CMAP terms were left unperturbed. We again stress that this perturbation only affects the dihedral potentials and that the free energy along the dihedral coordinate is largely influenced by nonbonded interactions between side chains.

Diffusion Coefficient. To estimate the diffusion coefficient along the end-to-end coordinate, we have used a method described earlier (5, 6). We consider a particle of reduced mass μ moving along a coordinate x, on the corresponding one-dimensional potential of mean force $W(x)$. Its motion follows a generalized Langevin equation (GLE):

$$\mu\ddot{x} = -\frac{\partial W(x)}{\partial x} - \int_0^t \zeta(t')\dot{x}(t-t')dt' + R(t), \qquad \text{[S2]}$$

where $\zeta(t)$ is the time-dependent friction and $R(t)$ is the random force that satisfies the fluctuation–dissipation theorem $\langle R(0)R(t)\rangle = k_{\mathrm{B}}T\zeta(t)$. The diffusion coefficient can be written as follows:

$$D = \frac{k_{\mathrm{B}}T}{\tilde{\zeta}(0)} = \frac{k_{\mathrm{B}}T}{\int_0^\infty \zeta(t)dt}, \qquad \text{[S3]}$$

where $\tilde{\zeta}(s)$ is the Laplace transform of $\zeta(t)$. In this approach, D is assumed to be position independent. For the present case of protein diffusion along its end-to-end distance L, D is expected to be spatially heterogeneous, i.e., $D(L)$. The following derivation aims to determine $D(L)$ at any position.

We first add a harmonic potential centered around a target value x_0 as follows:

$$W'(x) = \frac{1}{2}\mu\omega^2(x - x_0)^2, \qquad \text{[S4]}$$

where ω is the frequency of the biasing potential, and that we chose to be much stiffer than the actual PMF on which the protein is moving, so that the resulting PMF is locally harmonic around $\langle x \rangle \approx x_0$ with an effective frequency $\tilde{\omega}$. Using projection operators, one can recover the corresponding GLE as follows:

$$\mu\ddot{q} = -\mu\tilde{\omega}^2 q(t) - \int_0^t \zeta(t';x_0)\dot{x}(t-t')dt' + R(t), \qquad \text{[S5]}$$

with $q = x - \langle x \rangle$, $\mu\tilde{\omega}^2\langle q^2 \rangle = k_B T$. After multiplying by $q(0)$, taking the ensemble average and remembering that $\langle R(t)q(0) \rangle = 0$, the following expression is obtained for the position autocorrelation function $C_q(t)$:

$$\mu\ddot{C}_q(t) = -\mu\tilde{\omega}^2 C_q(t) - \int_0^t \zeta(t';x_0)\dot{C}_q(t-t')dt'. \qquad \text{[S6]}$$

We finally take the Laplace transform of this equation, leading after simplification to the following:

$$\mu\left(s^2\widetilde{C_q}(s) - s\langle q^2 \rangle\right) = -\mu\tilde{\omega}^2\widetilde{C_q}(s) - \tilde{\zeta}(s;x_0)\left(s\widetilde{C_q}(s) - \langle q^2 \rangle\right). \qquad \text{[S7]}$$

This equation can be written as follows:

$$\tilde{\zeta}(s;x_0) = \frac{\mu\tilde{\omega}^2\widetilde{C_q}(s)}{\langle q^2 \rangle - s\widetilde{C_q}(s)} - \mu s. \qquad \text{[S8]}$$

Combining Eq. **S3** and the limit $s \to 0$ of Eq. **S8**, and substituting L to x eventually leads to the following:

$$D(L \approx x_0) = \frac{\langle \delta L^2 \rangle^2}{\int_0^\infty \langle \delta L(0)\delta L(t) \rangle dt}, \qquad \text{[S9]}$$

where $\delta L = L - \langle L \rangle$ are the fluctuation of L around its average value.

At each given force, an additional 3-ns simulation in the microcanonical ensemble (time step of 1 fs) is performed using the collective-variable module of NAMD to add a bias potential on the end-to-end distance. This potential is harmonic and chosen to be much stiffer than the actual PMF on which the protein is moving. Here, we use a force constant of 100 kcal·mol^{-1}·Å^{-2} to constrain the system around a targeted end-to-end distance $\langle L_0 \rangle$. The unperturbed PMF is usually very smooth and we checked on a smaller, model system (decaalanine in water) using different values for the force constant (20, 50, and 100 kcal·mol^{-1}·Å^{-2}) that it does not have any significant impact on the value of D obtained. An example of the autocorrelation of L is shown in Fig. S8.

Determination of the Contour Length per Amino Acid. The fit of the force-extension data used in the manuscript can lead to uncertainties in the determination of the total contour length of the protein (i.e., its extension at infinite force). To minimize this effect, one can instead monitor the changes in length when adding or removing residues to the protein chain, as already suggested and used in an earlier experimental work (7). Here, we started from an extended configuration of ubiquitin at 250 and 100 pN from which we removed 10 and 20 residues, respectively. The resulting polypeptides were later equilibrated following the procedure described above. For each of them, we then performed a worm-like chain (WLC) fit of the end-to-end distance at the two forces using a fixed persistence length of 0.39 nm, leading to different contour lengths L_c shown in Fig. S1 as a function of the number of residues. We obtain a slope of 0.3815 nm in very good agreement with our earlier estimate of 0.38 nm, at the lower edge of the experimental estimate of 0.4 ± 0.02 nm (7) from force spectroscopy experiments. However, we achieve very good agreement between our value and the average distance between consecutive C_α of 0.38 nm found from protein structures in the PDB.

Effect of Force on Bonds and Bend Angles. The evolution of average backbone angles under force is shown in Fig. S3. In the force range used in most force spectroscopy experiments (<200–300 pN), very small changes are observed (typically less than 1%). A progressive increase is observed at higher forces, although this increase is very moderate even at forces around 1 nN. However, at much higher forces and thus extensions, description of bonds by classical force fields is expected to breakdown because the corresponding harmonic potentials cannot lead to bond rupture, which would require higher-level, quantum descriptions.

Distribution of Dihedral Angles Along the Sequence. As discussed in the main text, the average 1D and 2D distributions of dihedral angles at low forces may look similar to that of the folded protein; however, they are very different in nature, as illustrated in Fig. S4. For the folded state, the heterogeneous distribution arises from static heterogeneities among the amino acid sequence. For a given residue, the fluctuations of dihedral angles 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 because different local structures correspond to distinct values of these dihedral angles. Under force, where no secondary structure is observed, the individual fluctuations for each residue are very similar along the sequence with the notable exceptions of two of them, as detailed below. In this case, the average distribution is almost insensitive to the nature of the amino acid and therefore arises from dynamic disorder. A manuscript with further details and comparison between folded, chemically unfolded, and force-unfolded structures is currently under preparation.

Trajectories of Dihedral Angles After Force-Quench. We show in the main text that quenching the force results in the collapse of the end-to-end distance and in key changes in the Ramachandran plots. As discussed in detail, these two aspects are intimately connected. In fact, it is interesting to compare the simultaneous evolution of the average $|\psi|$ and $|\phi|$ angles (we consider the absolute values because the distance between the atoms 1 and 4 is independent of the sign of the angle) together with that of the end-to-end distance. Fig. S5 presents three such collapse trajectories when force is quenched from 250 pN down to 100 pN. The decrease in the end-to-end distance is very clearly correlated with that of one or both the dihedral angles. As a consequence, the observed diversity of these trajectories is seen to correspond to the various trajectories in the Ramachandran space.

Effect of Side Chains on Dihedral Angles. We have examined the variations of ϕ and ψ for each of the amino acids independently along the trajectories. In most cases, fluctuations are broad and poorly correlated with the size of the side chains. An illustration is given in Fig. S6 *A* and *B* for the ϕ angle of three adjacent amino acids (ILE13, THR14, LEU15) at two different forces. Although the shift to lower absolute values is observed when

force is lowered, the fluctuations of angles at a given force are quite similar.

Only two amino acids are found to behave very differently from the other ones, i.e., glycine and proline. The proline side chain is also involved in the backbone through a five-atom cycle, and it is therefore well known that it constrains dihedral orientations. For example, ϕ remains always close to −60° at any force (Fig. S6*D*). For glycine, which does not have any side chain, major differences are observed for the distributions of ϕ, which fluctuates around lower absolute values (Fig. S6*C*). Because ϕ can explore lower angular value more easily because of the absence of a side chain, the end-to-end C–C distance for each glycine is more sensitive to force, and therefore more flexible. The exploration of the Ramachandran space by the polyglycine analog of ubiquitin at any force is thus very different from that of the regular ubiquitin, as shown in Fig. S7. The greater flexibility of glycine is directly connected to this more spread-out exploration of the Ramachandran plot, which is consistent with the distributions obtained for glycine residues in the coil library of the PDB (8).

Alternate Determination of Diffusion Coefficients from Collapse Trajectories. It could be argued that, because of the stiff potentials used to determine the diffusion coefficients, we could miss small-amplitude but longer-timescale dynamics that would be relevant for the overall relaxation of the end-to-end distance in the absence of harmonic constraint. However, an alternative and approximate estimation of the diffusion coefficient from the collapse trajectories leads to very similar values. Although the average relaxation shown in Fig. 1*A* corresponding to a force quench from 250 to 100 pN is not monoexponential, we can extract a time constant of ∼0.5 ns by integrating the normalized decay. Because there is no analytical solution to the diffusion on a WLC potential, a second approximation is to consider that motion occurs on a harmonic potential, whose stiffness corresponds to the average root mean square displacement (RMSD) observed at 100 pN (∼0.3 nm). Diffusion in a 1D harmonic potential predicts an exponential relaxation with a time constant $\tau = \delta^2/D$, where δ is the RMSD and D is the diffusion coefficient. This leads to $D = (0.3)^2/0.5 = 0.18$ nm^2/ns $= 1.8 \times 10^8$ nm^2/s. This value only differs by a factor ∼2–3 compared with our original estimation of D (which is remarkable given the approximations introduced above and the simplicity of the model), showing that the timescales match.

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. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14(1):33–38.
5. Hummer G (2005) Position-dependent diffusion coefficients and free energies from bayesian analysis of equilibrium and replica molecular dynamics simulations. *New J Phys* 7:34.
6. Straub JE, Borkovec M, Berne BJ (1987) Calculation of dynamic friction on intramolecular degrees of freedom. *J Phys Chem* 91(19):4995–4998.
7. Ainavarapu SRK, et al. (2007) Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophys J* 92(1):225–233.
8. 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.

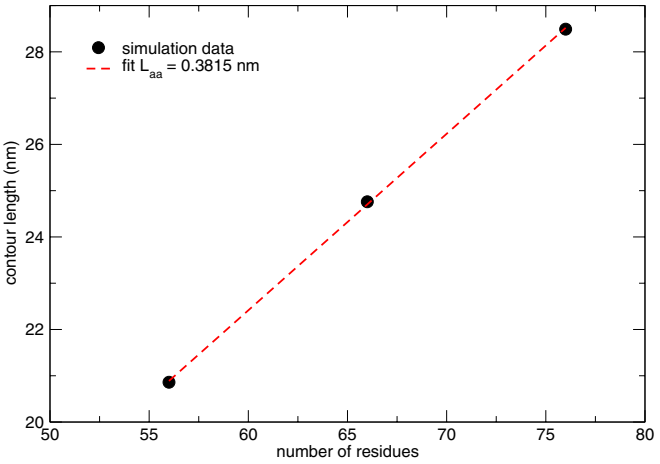

Fig. S1. Contour length as function of the number of residues for ubiquitin (76 residues) and its shortened configurations (66 and 56 residues, respectively). A linear fit leads to a peptide unit's length of 0.3815 nm.

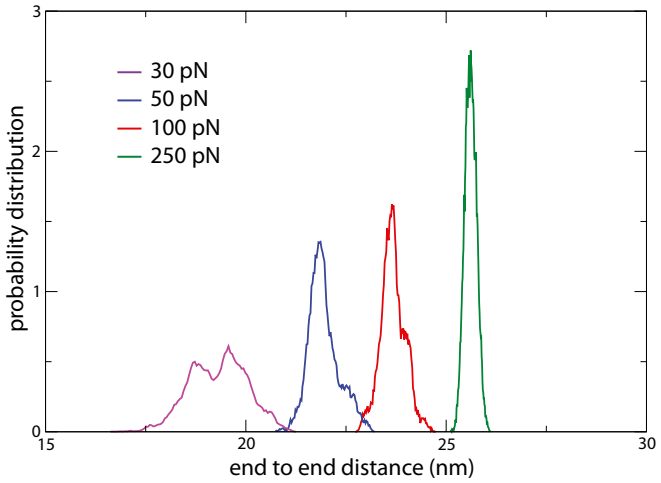

Fig. S2. Equilibrium distributions of the end-to-end length *L* at different forces.

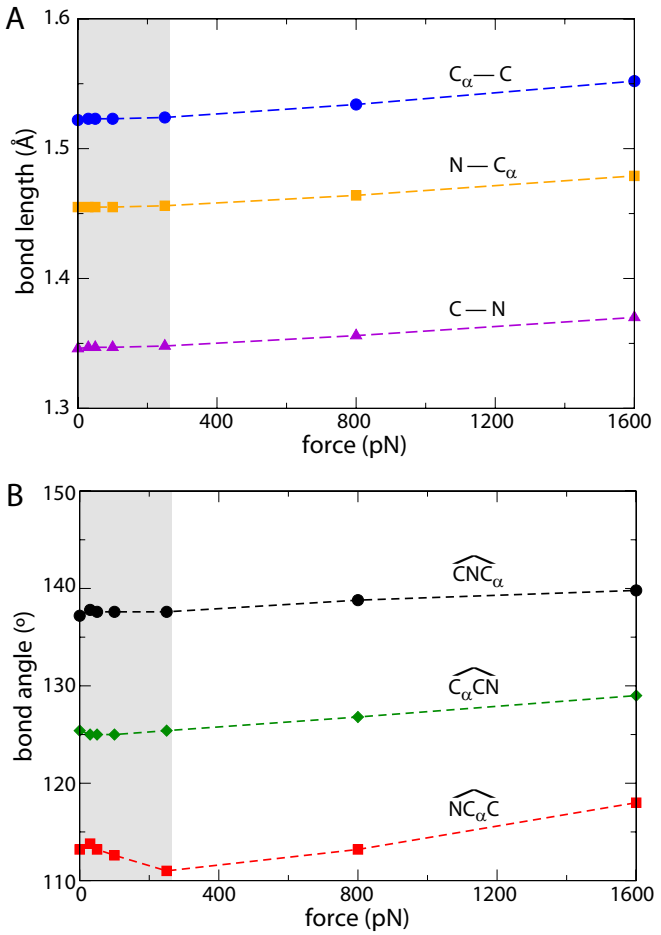

Fig. S3. Average backbone bonds (*A*) and angles (*B*) as a function of force (*A*: blue circles, C_α – C; orange squares, N – C_α; violet triangles, C – N; *B*: red squares, $\widehat{NC_\alpha C}$; green diamonds, $\widehat{C_\alpha CN}$; black circles, $\widehat{CNC_\alpha}$). The gray zone corresponds to the force range used in most force spectroscopy experiments.

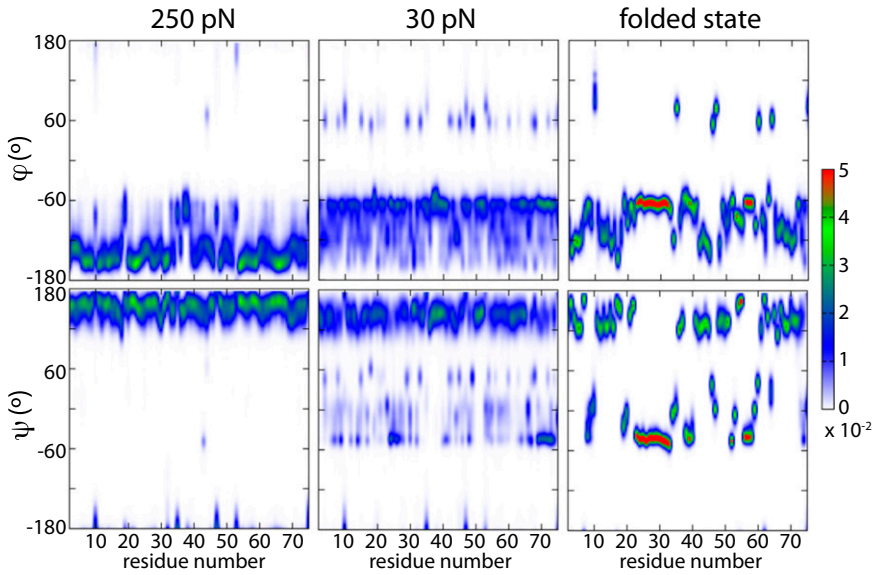

Fig. S4. Distribution of dihedral angles ϕ and ψ for each residue along the protein sequence for the unfolded protein at 250 and 30 pN together with that of the folded protein at zero force.

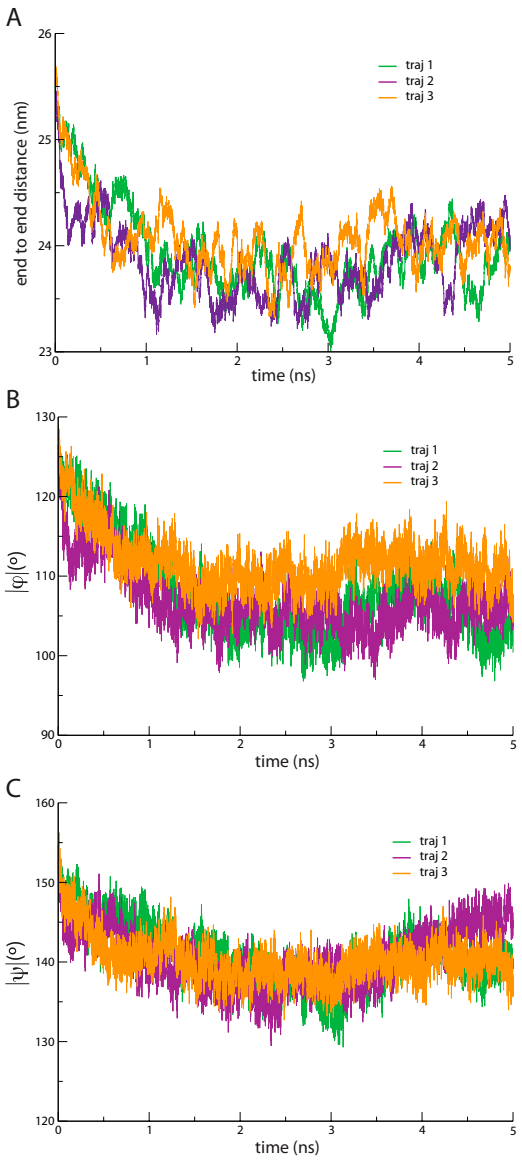

Fig. S5. Evolution of the end-to-end distance (*A*), the average $|\phi|$ (*B*), and $|\psi|$ (*C*) dihedral angles along three different trajectories where force has been quenched from 250 pN down to 100 pN at $t = 0$.

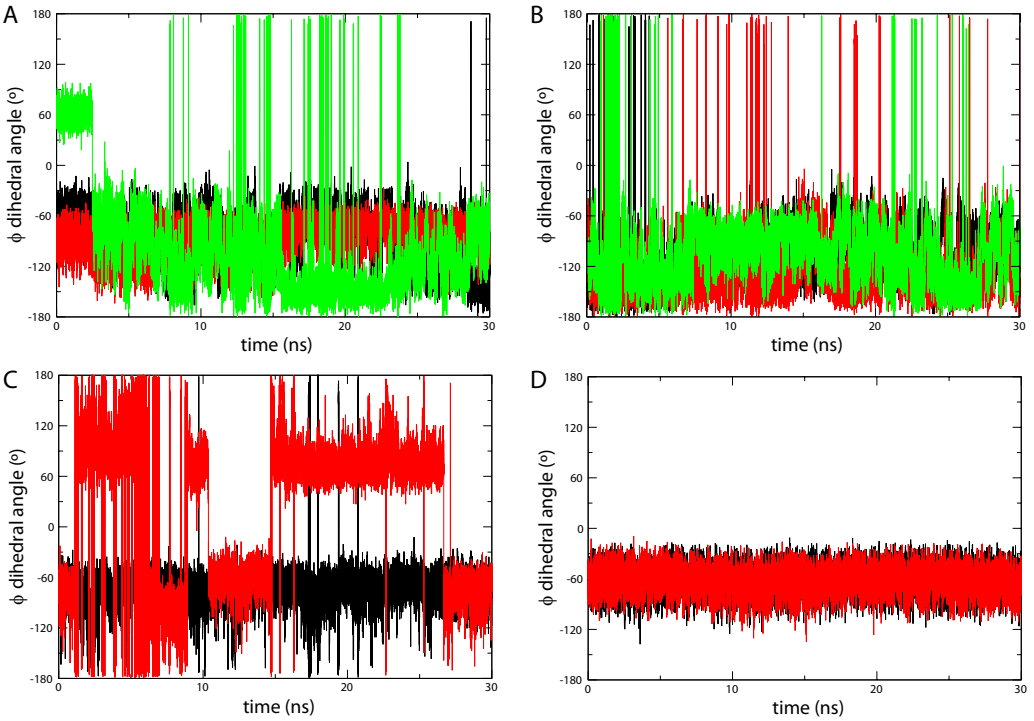

Fig. S6. Time evolution of ϕ for three consecutives residues (black, ILE13; red, THR14; green, LEU15) at (*A*) 30 pN and (*B*) 100 pN, (*C*) GLY10 (black) and GLY47 (red) at 100 pN, and (*D*) PRO19 (black) and PRO37 (red) at 100 pN.

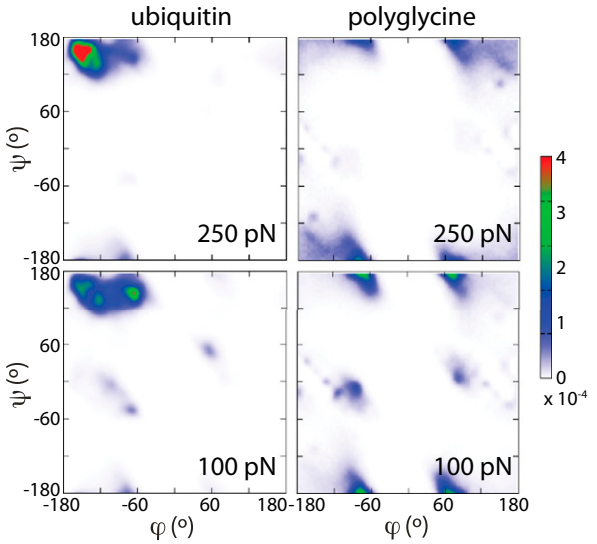

Fig. S7. Comparison between the Ramachandran plots of ubiquitin (*Left*) and its polyglycine analog (*Right*) at two different forces.

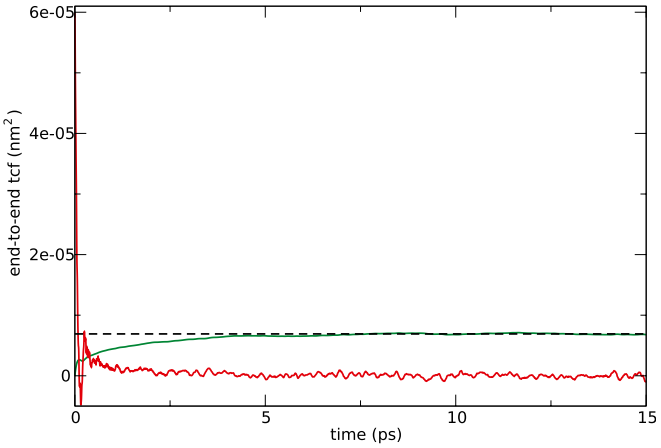

Fig. S8. Time correlation function of the normalized end-to-end length δL as a function of time (red curve) for ubiquitin at 100 pN. Its integral is shown in green, together with the extrapolated value used in Eq. **S9** (black dashes). The corresponding value of D is $(5.3 \pm 1.4)\cdot 10^8$ nm^2/s.