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 $\mathrm{kcal{\cdot}mol^{-1}{\cdot}\AA^{-2}}$ at the higher forces, a sampling every 2 Å with a force constant of 2.5 $\mathrm{kcal{\cdot}mol^{-1}{\cdot}\AA^{-2}}$ 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,

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

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):

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

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:

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

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:

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

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:

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

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 $C_q(t)$:

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

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

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

This equation can be written as follows:

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

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

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

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 $D = (0.3)^2/0.5 = 0.18\ \mathrm{nm}^2/\mathrm{ns} = 1.8 \times 10^8\ \mathrm{nm}^2/\mathrm{s}$. 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.

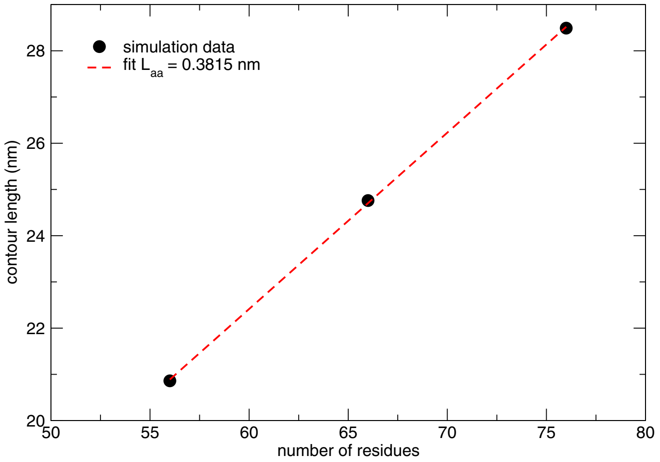

Fig. S1. Contour length as function of the number of residues for ubiquitin (76 residues) and its shortened configurations (66 and 56 residues, respectively). A linear fit leads to a peptide unit's length of 0.3815 nm.

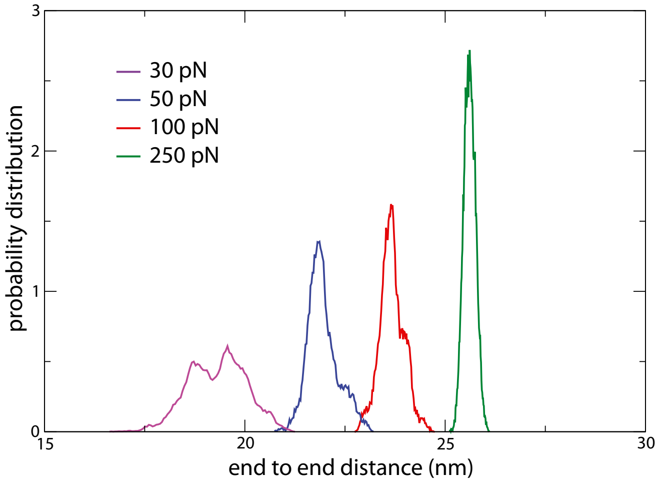

Fig. S2. Equilibrium distributions of the end-to-end length L at different forces.

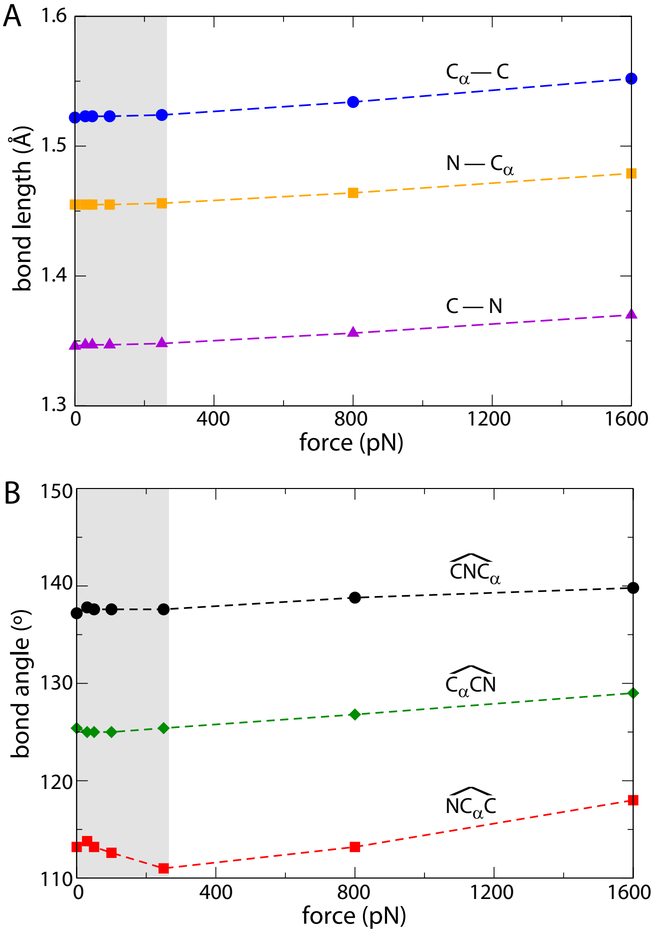

Fig. S3. Average backbone bonds (*A*) and angles (*B*) as a function of force (*A*: blue circles, $C_\alpha - C$; orange squares, $N - C_\alpha$; violet triangles, $C - N$; *B*: red squares, $\widehat{NC_\alpha C}$; green diamonds, $\widehat{C_\alpha CN}$; black circles, $\widehat{CNC_\alpha}$). The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

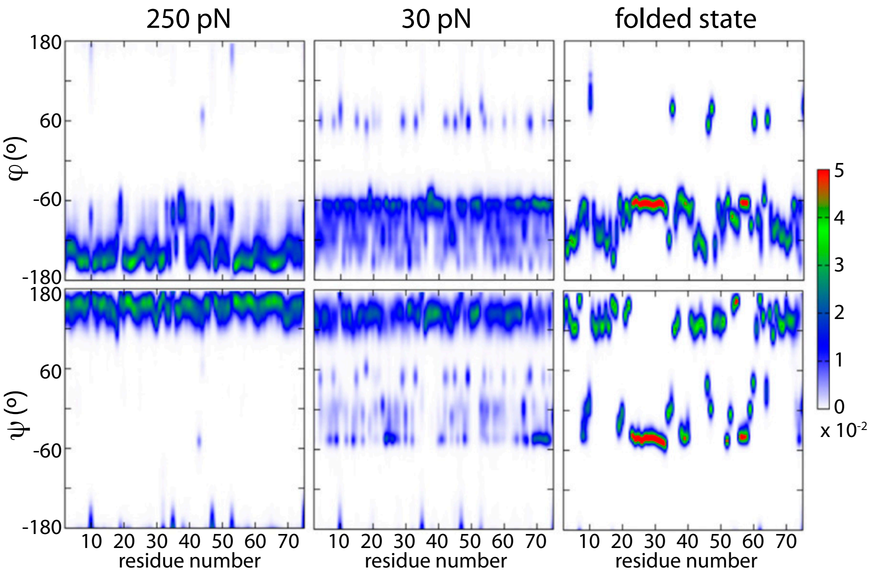

Fig. S4. Distribution of dihedral angles ϕ and ψ for each residue along the protein sequence for the unfolded protein at 250 and 30 pN together with that of the folded protein at zero force.

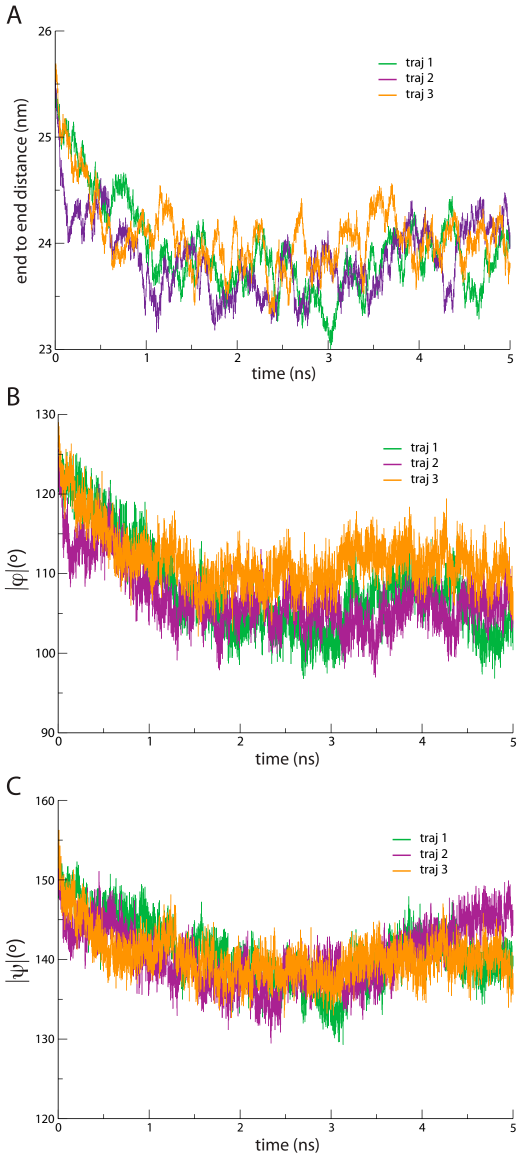

Fig. S5. Evolution of the end-to-end distance (*A*), the average $|\phi|$ (*B*), and $|\psi|$ (*C*) dihedral angles along three different trajectories where force has been quenched from 250 pN down to 100 pN at $t = 0$.

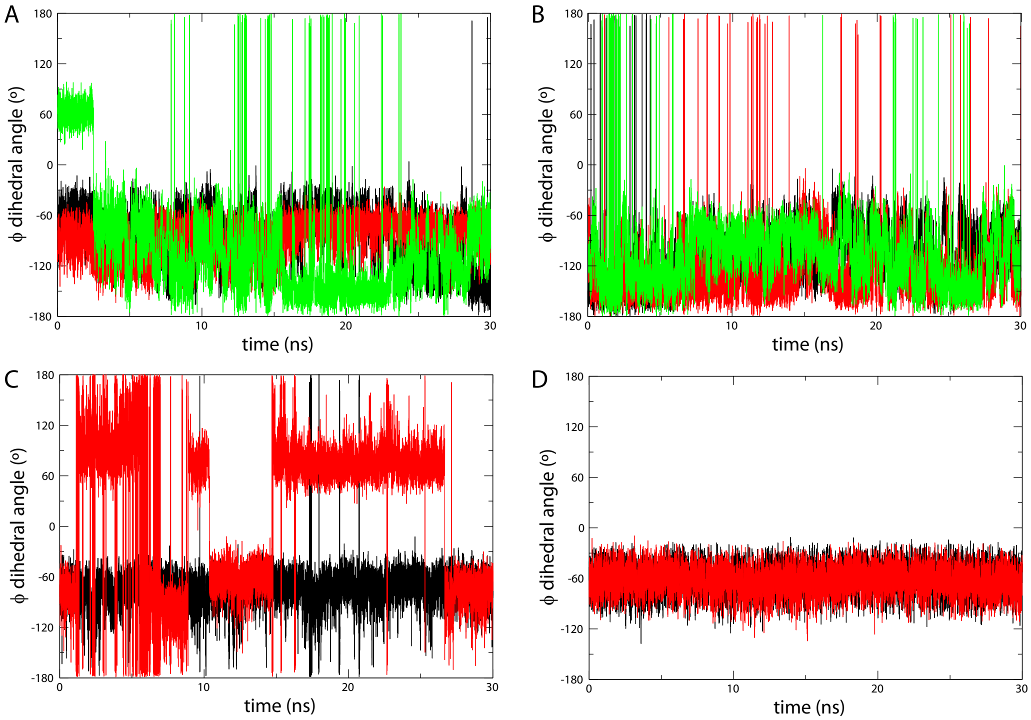

Fig. S6. Time evolution of ϕ for three consecutives residues (black, ILE13; red, THR14; green, LEU15) at (*A*) 30 pN and (*B*) 100 pN, (*C*) GLY10 (black) and GLY47 (red) at 100 pN, and (*D*) PRO19 (black) and PRO37 (red) at 100 pN.

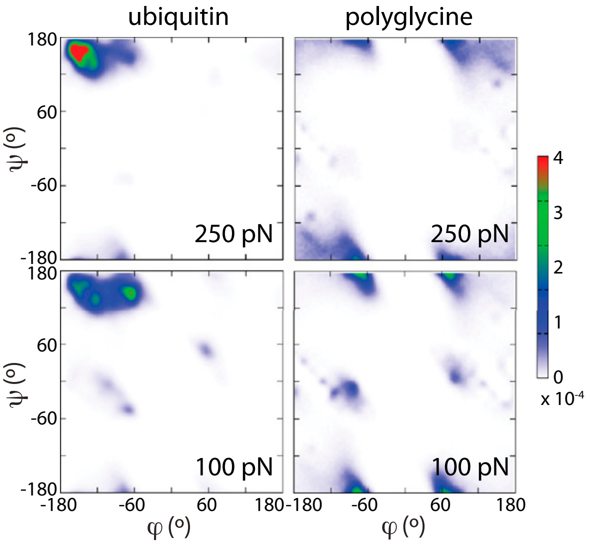

Fig. S7. Comparison between the Ramachandran plots of ubiquitin (*Left*) and its polyglycine analog (*Right*) at two different forces.

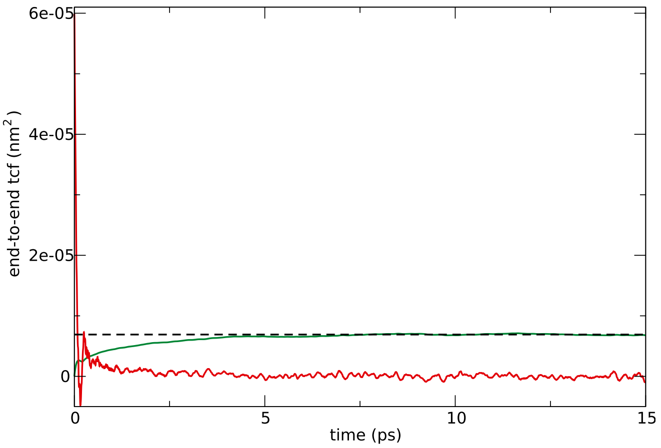

Fig. S8. Time correlation function of the normalized end-to-end length δL as a function of time (red curve) for ubiquitin at 100 pN. Its integral is shown in green, together with the extrapolated value used in Eq. **S9** (black dashes). The corresponding value of D is $(5.3 \pm 1.4)\cdot 10^8$ nm^2/s.