From mechanical folding trajectories to intrinsic energy landscapes of biopolymers

Michael Hinczewski[a,1], J. Christof M. Gebhardt[b,c], Matthias Rief[b], and D. Thirumalai[a,1]

[a]Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742; [b]Department of Physics, Technical University of Munich, 85748 Garching, Germany; and [c]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138

Edited* by Ken A. Dill, Stony Brook University, Stony Brook, NY, and approved January 23, 2013 (received for review August 13, 2012)

In single-molecule laser optical tweezer (LOT) pulling experiments, a protein or RNA is juxtaposed between DNA handles that are attached to beads in optical traps. The LOT generates folding trajectories under force in terms of time-dependent changes in the distance between the beads. How to construct the full intrinsic folding landscape (without the handles and beads) from the measured time series is a major unsolved problem. By using rigorous theoretical methods—which account for fluctuations of the DNA handles, rotation of the optical beads, variations in applied tension due to finite trap stiffness, as well as environmental noise and limited bandwidth of the apparatus—we provide a tractable method to derive intrinsic free-energy profiles. We validate the method by showing that the exactly calculable intrinsic free-energy profile for a generalized Rouse model, which mimics the two-state behavior in nucleic acid hairpins, can be accurately extracted from simulated time series in a LOT setup regardless of the stiffness of the handles. We next apply the approach to trajectories from coarse-grained LOT molecular simulations of a coiled-coil protein based on the GCN4 leucine zipper and obtain a free-energy landscape that is in quantitative agreement with simulations performed without the beads and handles. Finally, we extract the intrinsic free-energy landscape from experimental LOT measurements for the leucine zipper.

The energy landscape perspective has provided a conceptual framework to describe how RNA (1) and proteins (2–4) fold. Some of the key theoretical predictions (5, 6), have been confirmed by experiments (7). More refined comparisons require mapping the full folding landscape of biomolecules. Advances in laser optical tweezer (LOT) experiments have been used to obtain free-energy profiles as a function of the extension of biomolecules under tension (7–12).

The usefulness of the LOT technique, however, hinges on the assumption that information about the fluctuating biomolecule can be accurately recovered from the raw experimental data, namely the time-dependent changes in the positions of the beads in the optical traps, attached to the biomolecule by double-stranded DNA (dsDNA) handles (Fig. 1). Thus, we only have access to the intrinsic folding landscape of the biomolecule (in the absence of handles and beads) indirectly through the bead–bead separation along the force direction. Many extraneous factors, such as fluctuations of the handles (13, 14), rotation of the beads, and the varying applied tension due to finite trap stiffness, can distort the intrinsic folding landscape. Moreover, the detectors and electronic systems used in the data collection have finite response times, leading to filtering of high-frequency components in the signal (15). Ad hoc attempts have been made to account for handle effects based on experimental estimates of stretched DNA properties, using techniques similar to image deconvolution (8, 11, 16). Theory has been used to extract free-energy information from nonequilibrium pulling experiments (17), and to determine the intrinsic power spectrum of protein fluctuations (18) from LOT data. However, to date there has been no comprehensive theory to model and correct for all of the systematic instrumental distortions of the underlying folding landscapes of proteins and RNA.

How can one construct the intrinsic free-energy profile of a biomolecule using the measured folding trajectories in the presence of beads and handles [the total separation $z_{\rm tot}(t)$ in Fig. 1 as a function of time t]? Here, we solve this problem using a rigorous theoretical procedure. Besides $z_{\rm tot}(t)$, the only inputs needed in our theory are the bead radii, the trap strengths and positions, and handle characteristics such as the contour length, the persistence length, and the elastic stretch modulus. The output is the intrinsic free energy as a function of the biomolecular extension ($z_{\rm p}$ in Fig. 1) in the constant force ensemble.

We validate our approach using two systems: (*i*) a generalized Rouse model (GRM) hairpin (19), which has an analytically solvable double-well energy landscape under force; and (*ii*) a double-stranded coiled-coil protein based on the yeast transcriptional factor GCN4 leucine zipper domain, whose folding landscape was studied using a LOT experiment (11). We first use coarse-grained molecular simulations to obtain the intrinsic free-energy landscape of the isolated protein at a constant force. We then simulate mechanical folding trajectories using the full LOT setup, from which we quantitatively recover the intrinsic free-energy landscape of GCN4, thus further establishing the efficacy of our theory. Finally, we apply our theory to experimentally generated data, and show that we can get reliable estimates for the protein energy profile independent of the optical trap parameters.

Results

Theory for Constructing the Intrinsic Protein Folding Landscape from Measurements. In a dual-beam optical tweezer setup (Fig. 1) the protein is covalently connected to dsDNA handles that are attached to glass or polystyrene beads in two optical traps. For small displacements of the beads from the trap centers (20) the trap potentials are harmonic, with strengths $k_x = k_z \equiv k_{\rm trap}$ along the lateral plane, and a weaker axial strength $k_y = \alpha k_{\rm trap}$, where $\alpha < 1$ (21). For simplicity, we take both traps to have equal strengths, although our method can be generalized to an asymmetric setup. The trap centers are separated from each other along the $\hat{\mathbf{z}}$ axis, with trap 1 at $z = 0$ and trap 2 at $z = z_{\rm trap}$. As the bead–handle–protein system fluctuates in equilibrium, the positions of the bead centers $\mathbf{r}_1(t)$ and $\mathbf{r}_2(t)$ vary in time. The experimentalist can collect a time series of the z components of the bead positions $z_1(t)$ and $z_2(t)$. Denote the mean of each time series as $\bar{z}_1$ and $\bar{z}_2$. We assume that the trap centers are sufficiently far apart that the whole system is under tension, which implies that the mean bead displacements are nonzero, $\bar{z}_1 = z_{\rm trap} - \bar{z}_2 = \overline{F}/k_{\rm trap} > 0$, where $\overline{F}$ is the mean tension along $\hat{\mathbf{z}}$. We focus on the case where there is no feedback mechanism to maintain a constant force, so the instantaneous tension in the system changes as the total end-to-end extension component $z_{\rm tot}(t) \equiv z_2(t) - z_1(t)$ (Fig. 1) varies. Although we choose one particular passive setup, the theory can be adapted to other types of passive optical tweezer systems (8, 20) where the force is

Author contributions: M.H. and D.T. designed research; M.H. and J.C.M.G. performed research; M.H., M.R., and D.T. analyzed data; and M.H., M.R., and D.T. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence may be addressed. E-mail: mhincz@umd.edu or thirum@umd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1214051110/-/DCSupplemental.

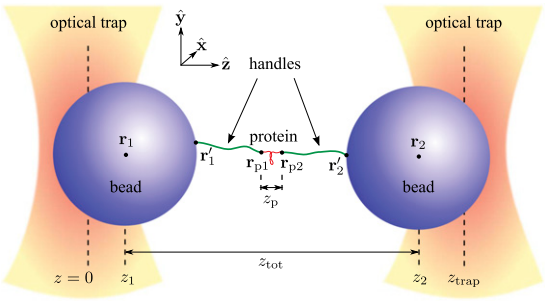

Fig. 1. Dual-beam optical tweezer setup for studying the equilibrium folding landscape of a single protein molecule under force.

approximately constant (in which case we could skip the transformation into the constant-force ensemble described below). The mean tension $\overline{F}$, a measure of the overall force scale, can be tuned at the start of the experiment by making the trap separation z_{trap} larger (leading to higher $\overline{F}$) or smaller (leading to lower $\overline{F}$). Because $\overline{F} = k_{trap}(z_{trap} - \overline{z}_{tot})/2$, the precise relationship between z_{trap} and $\overline{F}$ requires knowing the mean total extension $\overline{z}_{tot}$, which depends among other things on the details of the energy landscape. Hence, we cannot in general calculate beforehand what $\overline{F}$ will be for a given z_{trap}. However, one of the advantages of our approach is that we can combine data from different experimental runs (each having a different z_{trap} and $\overline{F}$) to accurately construct the protein free-energy profile. This combination is carried out through the weighted histogram analysis method (WHAM) (22) (*Supporting Information*) in a spirit similar to earlier work in the context of optical tweezers (23, 24). We first solve the problem of obtaining the protein landscape based on a single observed trajectory of bead-to-bead separations specified as z_{tot} as a function of t.

The key quantity in the construction procedure is $\mathcal{P}_{tot}(z_{tot})$, the equilibrium probability distribution of z_{tot} within the external trap potential, which can be directly derived from the experimental time series. The imperfect nature of the measured data, due to noise and low-pass filtering effects in the recording apparatus, will distort $\mathcal{P}_{tot}(z_{tot})$, but we have developed a technique to model and approximately correct for these issues (*Materials and Methods*, *FBS*). Once we have an experimental estimate for $\mathcal{P}_{tot}(z_{tot})$, the objective is to find $\tilde{\mathcal{P}}_p(z_p; F_0)$, the intrinsic distribution of the protein end-to-end extension component z_p at some *constant* force F_0, whose value we are free to choose. (Tilde notation denotes probabilities in the constant-force ensemble.) The intrinsic protein free-energy profile is $\tilde{\mathcal{F}}_p(z_p; F_0) = -k_B T \ln \tilde{\mathcal{P}}_p(z_p; F_0)$. The procedure, obtained from rigorous theoretical underpinnings described in detail in *SI Text*, consists of two steps:

i) Transformation into the constant-force ensemble. Given $\mathcal{P}_{tot}(z_{tot})$, we obtain the total system end-to-end distribution at a constant F_0 using

$$\tilde{\mathcal{P}}_{tot}(z_{tot}; F_0) = C^{-1} e^{\beta F_0 z_{tot} + \frac{1}{4}\beta k_{trap}\left(z_{trap} - z_{tot}\right)^2} \mathcal{P}_{tot}(z_{tot}), \qquad [1]$$

where $\beta = 1/k_B T$ and C is a normalization constant. The equation above applies in the case of a single experimental trajectory at a particular trap separation z_{trap}.

ii) Extraction of the intrinsic protein distribution. In the constant-force ensemble, $\tilde{\mathcal{P}}_{tot} = \tilde{\mathcal{P}}_b * \tilde{\mathcal{P}}_h * \tilde{\mathcal{P}}_p * \tilde{\mathcal{P}}_h * \tilde{\mathcal{P}}_b$, relates the total end-to-end fluctuations $\tilde{\mathcal{P}}_{tot}(z_{tot}; F_0)$ to the end-to-end distributions for the individual components $\tilde{\mathcal{P}}_\alpha(z_\alpha; F_0)$, where α denotes bead (b), handle (h), or protein (p), and "*" is a 1D convolution operator. For the beads, "end-to-end" refers to the extension between the bead center and the handle attachment point, projected along $\hat{\mathbf{z}}$. In Fourier space the convolution has the form

$$\begin{aligned}\tilde{\mathcal{P}}_{tot}(k; F_0) &= \tilde{\mathcal{P}}_b^2(k; F_0)\tilde{\mathcal{P}}_h^2(k; F_0)\tilde{\mathcal{P}}_p(k; F_0) \\ &\equiv \tilde{\mathcal{P}}_{bh}(k; F_0)\tilde{\mathcal{P}}_p(k; F_0),\end{aligned} \qquad [2]$$

where $\tilde{\mathcal{P}}_\alpha(k; F_0)$ is the Fourier transform of $\tilde{\mathcal{P}}_\alpha(z_\alpha; F_0)$. Here, $\tilde{\mathcal{P}}_{bh}$, which is the result of convolving all of the bead and handle distributions, acts as the main point-spread function relating the intrinsic protein distribution $\tilde{\mathcal{P}}_p$ to $\tilde{\mathcal{P}}_{tot}$. Because $\tilde{\mathcal{P}}_{bh}$ can be modeled from a theoretical description of the handles and beads, we can solve for $\tilde{\mathcal{P}}_p$ using Eq. **2** and hence find $\tilde{\mathcal{F}}_p$, the intrinsic free-energy profile of the protein.

The derivation of the procedure (*SI Text*) shows the conditions under which the two-step method works. The mathematical approximation underlying step *i* becomes exact if: (*i*) $k_x = k_y = 0$; or (*ii*) the full 3D total system end-to-end probability is separable into a product of distributions for longitudinal ($\hat{\mathbf{z}}$) and transverse ($\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$) components. In general, condition (*ii*) is not physically sensible (19). However, if $\overline{\rho}_{tot}$ is the typical length scale describing transverse fluctuations, then condition (*i*) is approximately valid when $\beta k_{trap}\overline{\rho}_{tot}^2 \ll 1$. If this condition breaks down, accurate construction of the intrinsic energy landscape cannot be performed without knowledge of the transverse behavior. In the simulation and experimental results below, the force scales are such that transverse fluctuations are small, $\overline{\rho}_{tot} \sim \mathcal{O}(1\ \text{nm})$, so to ensure condition (*i*) is met, we require that $k_{trap} \ll k_B T/\overline{\rho}_{tot}^2 = 4.1$ pN/nm at $T = 298$ K. We use the experimental value $k_{trap} = 0.25$ pN/nm in our test cases (11), which is well under the upper limit. In principle, one can choose any F_0, the force value of the constant-force ensemble where we carry out the analysis. In practice, F_0 should be chosen from among the range of forces that is sampled in equilibrium during the actual experiment, because this will minimize statistical errors in the final constructed landscape. For example, setting $F_0 = \overline{F}$, the mean tension, is a reasonable choice.

Step *ii* depends on knowledge of $\tilde{\mathcal{P}}_{bh}(k; F_0)$, and thus the individual constant-force distributions of the beads and the handles in Fourier space. The point-spread function is characterized by the bead radius R_b, the handle contour length L, the handle persistence length l_p, and the handle elastic stretching modulus γ. In $\tilde{\mathcal{P}}_h$ we also include the covalent linkers which attach the handles to the beads and protein, giving two additional parameters: the linker stiffness κ and length ℓ. Using the extensible semiflexible chain as a model for the handles, we exploit an exact mapping between this model and the propagator for the motion of a quantum particle on the surface of a unit sphere (25) to calculate the handle Fourier-space distribution to arbitrary numerical precision. Together with analytical results for the bead and linker distributions, we can directly solve for $\tilde{\mathcal{P}}_{bh}(k; F_0)$. To verify that the analytical model for the point-spread function can accurately describe handle/bead fluctuations over a range of forces, we have analyzed data from control experiments on a system involving only the dsDNA handles attached to beads, where $\mathcal{P}_{tot} = \mathcal{P}_{bh}$ (*SI Text*). The theory simultaneously fits results for several experimental quantities measured on the same system: the distributions $\tilde{\mathcal{P}}_{bh}$ derived from three different trap separations, corresponding to mean forces F_0 = 9.4–12.7 pN, and a force-extension curve. The accuracy of the model $\tilde{\mathcal{P}}_{bh}$ is ~1–3%, within the experimental error margins.

Robustness of the Theory Validated by Application to an Exactly Solvable Model. We first apply the theory to a problem for which the intrinsic free-energy profiles at arbitrary force are known exactly. The GRM hairpin (*SI Text*) is a two-state folder whose full 3D equilibrium end-to-end distributions are analytically solvable. A representative GRM distribution $\tilde{\mathcal{P}}_{GRM}$ at F_0 = 11.9 pN is plotted in Fig. 2*A*. The upper part shows a projection onto the $(\rho = \sqrt{x^2 + y^2}, z)$ plane, because $\tilde{\mathcal{P}}_{GRM}$ is cylindrically symmetric, whereas the lower part shows the further projection onto the z coordinate. The two peaks correspond to the native

(N) state at small z, and the unfolded (U) state at large z. To model the optical tweezer system, we add handles and beads to the GRM hairpin, whose probabilities $\tilde{\mathcal{P}}_{\rm h}$ and $\tilde{\mathcal{P}}_{\rm b}$ (including transverse fluctuations) are illustrated in Fig. 2 *B* and *C*. The full 3D behavior is derived in an analogous manner to the theory mentioned above for the 1D Fourier-space distribution $\tilde{\mathcal{P}}_{\rm bh}(k;F_0)$ of the beads/handles; the only difference is that the transverse degrees of freedom are not integrated out. The 3D convolution of the system components, plus the optical trap contribution, gives the total distribution $\mathcal{P}_{\rm tot}$ in Fig. 2*D*. The bead, handle, linker, and trap parameters are listed in Table S1. From $\mathcal{P}_{\rm tot}$ one can calculate the mean total z extension and the mean tension, which in this case are $\overline{z}_{\rm tot}=1,199$ nm, $\overline{F}=k_{\rm trap}(z_{\rm trap}-\overline{z}_{\rm tot})/2=11.9$ pN.

The $\hat{\mathbf{z}}$-probability projection in Fig. 2*D* (*Lower*) is the information accessible in an experiment, and the computation of the intrinsic distribution in Fig. 2*A* (*Lower*) is the ultimate goal of the construction procedure. Comparing *A* and *D*, two effects of the apparatus are visible: the GRM peaks have been partially blurred into each other, and the transverse (ρ) fluctuations have been enhanced. The handles provide the dominant contribution to both these effects.

Fig. 2 *E*–*G* illustrates the construction procedure for the GRM optical tweezer system. *E* corresponds to step *i*, with a transformation of the distribution $\mathcal{P}_{\rm tot}$ (whose varying force scale is shown along the top axis) into $\tilde{\mathcal{P}}_{\rm tot}$ at constant force F_0 = 11.9 pN. Step *ii* uses the exact $\tilde{\mathcal{P}}_{\rm bh}$, shown in real space in *F*, and produces the intrinsic distribution $\tilde{\mathcal{P}}_{\rm GRM}$, drawn as a solid line in *G*. The agreement with the exact analytical result (dashed line) is extremely close, with a median error of 3% over the range shown. This deviation is due to the approximation in step *i*, discussed above, as well as the numerical implementation of the deconvolution procedure.

As shown in our previous study (19), the smaller the ratio l_p/L for the handles, the more the features of the protein energy landscape get blurred by the handle fluctuations. Because the experimentally measured total distribution always distorts to some extent the intrinsic protein free-energy profile due to the finite duration and sampling of the system trajectory, more flexible handles will exacerbate the signal-to-noise problem. To illustrate this effect, we performed Brownian dynamics simulations of the GRM in the optical tweezer setup, with handles modeled as extensible, semiflexible bead–spring chains (*SI Text*). In Fig. 3*A* we compare the free energy $\mathcal{F}_{\rm tot}=-k_B\,T\ln\mathcal{P}_{\rm tot}$ for a fixed L = 100 nm and a varying l_p/L, derived from the simulation trajectories, and the exact intrinsic GRM result $\tilde{\mathcal{F}}_{\rm GRM}=-k_BT\ln\tilde{\mathcal{P}}_{\rm GRM}$ at F_0. When the handles are very flexible, with l_p/L = 0.02, the energy barrier between the native and unfolded states almost entirely disappears in $\mathcal{F}_{\rm tot}$, with the noise making the precise barrier shape difficult to resolve. Remarkably, even with this extreme level of distortion, using our theory we still recover a reasonable estimate of the intrinsic landscape (Fig. 3*B*). For each $\mathcal{F}_{\rm tot}$ in Fig. 3*A*, Fig. 3*B* compares the result of the construction procedure and the exact answer for $\tilde{\mathcal{F}}_{\rm GRM}$. Clearly some information is lost as l_p/L becomes smaller, because the l_p/L = 0.02 system does not yield as accurate a result as the ones with stiffer handles. However, in all cases the basic features of the exact $\tilde{\mathcal{F}}_{\rm GRM}$ are reproduced. Thus, the method works remarkably well over a wide range of handle parameters. This conclusion is generally valid even when other parameters are varied (see Fig. S3 in *SI Text* for tests at various F_0 and $k_{\rm trap}$). The excellent agreement between the constructed and intrinsic free-energy profiles for the exactly solvable GRM hairpin over a wide range of handle and trap experimental variables establishes the robustness of the theory.

Intrinsic Folding Landscape of a Simulated Leucine Zipper. To demonstrate that the theory can be used to produce equilibrium intrinsic free-energy profiles with multiple states from mechanical folding trajectories, we performed simulations of a protein in an optical tweezer setup. The simulations were designed to mirror a single-molecule experiment (11). To this end we studied a coiled-coil, LZ26 (26), based on three repeats of the leucine zipper domain from the yeast transcriptional factor GCN4 (27) (*Materials and Methods*). The simple linear unzipping of the two strands of LZ26 allows us to map the end-to-end extension to the protein configuration. Furthermore, the energy heterogeneity of the native bonds that form the "teeth" of the zipper leads to a nontrivial folding landscape with at least two intermediate states (11, 26, 28).

The LZ26 N structure in Fig. 4 (from a simulation snapshot), shows two alpha-helical strands running from the N terminus at the bottom to the C terminus at the top. In the experiment a handle is attached to the N terminus of each strand, and this is where the strands begin to unzip under applied force. To prevent complete strand separation, the C termini are cross-linked through a disulfide bridge between two cysteine residues. Each alpha-helix coil consists of a series of seven-residue heptad repeats, with positions labeled a through g. For the leucine zipper the a and d positions are the teeth, consisting of mostly hydrophobic residues (valine and leucine) which have strong noncovalent interactions with their counterparts on the other strand. The exceptions to the hydrophobic pattern are the three hydrophilic asparagine residues in a position on each strand

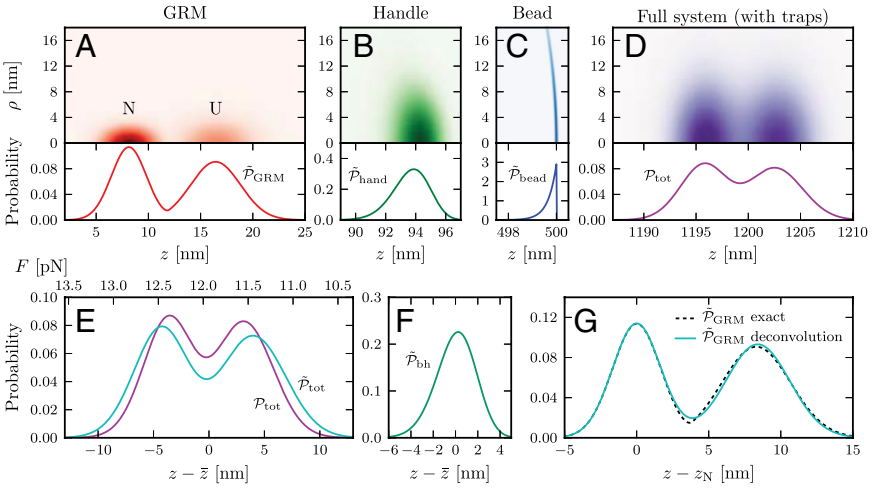

Fig. 2. GRM hairpin in an optical tweezer setup. First row shows the exact end-to-end distributions along $\hat{\mathbf{z}}$ for each component type in the system: (*A*) GRM, (*B*) dsDNA handle, (*C*) polystyrene bead. Handle, bead, and trap parameters are listed in Table S1 (GRM column). (*Upper*) Probabilities projected onto cylindrical coordinates ($\rho=\sqrt{x^2+y^2}, z$). (*Lower*) Projection onto z alone. (*D*) Result for the total system end-to-end distribution $\mathcal{P}_{\rm tot}$ derived by convolving the component probabilities and accounting for the optical traps. (*E*–*G*) Construction of the original GRM distribution $\tilde{\mathcal{P}}_{\rm GRM}$ starting from $\mathcal{P}_{\rm tot}$. (*E*) $\mathcal{P}_{\rm tot}$ (purple) and $\tilde{\mathcal{P}}_{\rm tot}$ (blue) as a function of z on the bottom axis, measured relative to $\overline{z}$, the average extension for each distribution. For $\mathcal{P}_{\rm tot}$, the upper axis shows the z range translated into the corresponding trap forces F. After removing the trap effects, $\tilde{\mathcal{P}}_{\rm tot}$ is the distribution for constant force F_0 = 11.9 pN. (*F*) $\tilde{\mathcal{P}}_{\rm bh}$, describing the total probability at F_0 of fluctuations resulting from both handles and the rotation of the beads. (*G*) Constructed solution for $\tilde{\mathcal{P}}_{\rm GRM}$ (solid line), obtained by numerically inverting the convolution $\tilde{\mathcal{P}}_{\rm tot}=\tilde{\mathcal{P}}_{\rm bh}*\tilde{\mathcal{P}}_{\rm GRM}$. Exact analytical result for $\tilde{\mathcal{P}}_{\rm GRM}$ is shown as a dashed line; $z_{\rm N}$ is the position of the N peak.

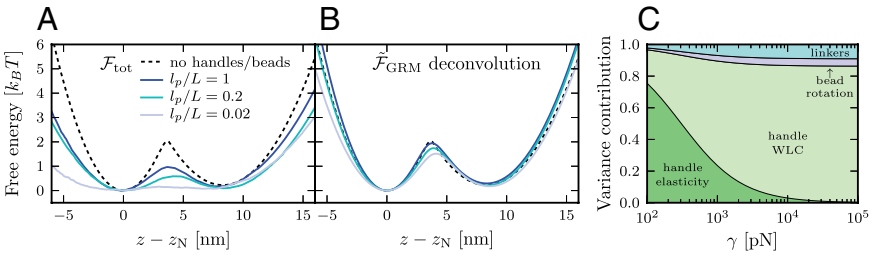

Fig. 3. Effects of handle characteristics on the free-energy profile of the GRM in a LOT setup. (*A*) Total system free energy $\mathcal{F}_{\rm tot} = -k_BT \ln \mathcal{P}_{\rm tot}$ for fixed L = 100 nm, and varying ratios l_p/L. All of the other parameters are in Table S1 (GRM column). The exact analytical free energy at F_0 =11.9 pN (dashed line) for the GRM alone, $\tilde{\mathcal{F}}_{\rm GRM} = -k_BT \ln \tilde{\mathcal{P}}_{\rm GRM}$, is shown for comparison. (*B*) For each $\mathcal{F}_{\rm tot}$ in *A*, the construction of $\tilde{\mathcal{F}}_{\rm GRM}$ at F_0, together with the exact answer (dashed line). (*C*) For system parameters matching the experiment (Table S1), the variance of the point-spread function $\tilde{\mathcal{P}}_{\rm bh}$ broken down into the individual handle, bead, and linker contributions. The fraction for each component is shown as a function of varying handle elastic modulus γ.

(marked in blue in the structure snapshots in Fig. 4). As has been seen experimentally (11, 26) (and shown below through simulations), the weaker interaction of these asparagine pairs is crucial in determining the properties of the intermediate folding states.

In analyzing the LZ26 leucine zipper system, we performed coarse-grained simulations using the self-organized polymer (SOP) model (29) (full details in *SI Text*, with parameters summarized in Table S1). The intrinsic free-energy profile $\tilde{\mathcal{F}}_{\rm p} = -k_BT \ln \tilde{\mathcal{P}}_{\rm p}$ at $F_0 = 12.3$ pN in Fig. 4*A* has four prominent wells in $\tilde{\mathcal{F}}_{\rm p}$ as a function of $z_{\rm p}$ corresponding to four stages in the progressive unzipping of LZ26. At $F_0 = 12.3$ pN all of the states are populated, and the system fluctuates in equilibrium between the wells. The transition barrier between N and I1 exhibits a shallow dip that may correspond to an additional, very transiently populated intermediate. Because this dip is much smaller than k_BT, we do not count it as a distinct state.

Adding the optical tweezer apparatus to the SOP simulation significantly distorts the measured probability distributions. In the first row of Fig. 5 sample simulation trajectory fragments are shown both for the protein-only case (Fig. 5*A*) at constant force $F_0 = 12.3$ pN, and within the full optical tweezer system (Fig. 5*C*) with $z_{\rm trap} = 503$ nm. For the latter case we plot both $z_{\rm tot}(t)$ (purple) and $z_p(t)$ (gray), allowing us to see how the bead separation tracks changes in the protein extension. The probability distributions $\tilde{\mathcal{P}}_{\rm p}$ and $\mathcal{P}_{\rm tot}$ are plotted in Fig. 5 *B* and *D*, respectively. In Fig. 5*E*, the distribution $\mathcal{P}_{\rm tot}$ within the optical tweezer system is plotted for $z_{\rm trap} = 503$ nm. Although we only illustrate this particular $z_{\rm trap}$ value, ~260 trajectories are generated at different $z_{\rm trap}$ and combined together using WHAM (22) (*SI Text*) to produce a single $\tilde{\mathcal{P}}_{\rm tot}$ at a constant force $F_0 = 12.3$ pN (Fig. 5*E*). We can then use our theoretical method to recover the protein free energy $\tilde{\mathcal{F}}_{\rm p}$ (Fig. 5*F*). Despite numerical errors due to limited statistical sampling (both in the protein-only and total system runs), there is remarkable agreement between the constructed result and $\tilde{\mathcal{F}}_{\rm p}$ derived from protein-only simulations. This is particularly striking given that the total system free energy $\mathcal{F}_{\rm tot}(z_{\rm tot}) = -k_BT \ln \mathcal{P}_{\rm tot}(z_{\rm tot})$ (Fig. 5*F*), shows that handles/beads blur the energy landscape, reducing the energy barriers to a degree that the N state is difficult to resolve. The signature of N in $\mathcal{F}_{\rm tot}(z_{\rm tot})$ is a slight change in the curvature at higher energies on the left of the I1 well. Nevertheless, we still recover a basin of attraction representing the N state in the constructed $\tilde{\mathcal{F}}_{\rm p}$. The results in Fig. 5 provide a self-consistency check of the method for a system with multiple intermediates.

Folding Landscape of the Leucine Zipper from Experimental Trajectories. As a final test of the efficacy of the theory we used the experimental time series data (11) to obtain $\tilde{\mathcal{F}}_{\rm p}$. The data consist of two independent runs with the LZ26 leucine zipper, using the same handle/bead parameters for each run (Table S1) but at different trap separations $z_{\rm trap}$. We project the deconvolved landscape from each trajectory onto the midpoint force F_0 where the two most populated states (I1 and U) have equal probabilities in $\tilde{\mathcal{P}}_{\rm p}$. The values of F_0 derived from the two runs are the same within error bounds: 12.3 ± 0.9 and 12.1 ± 0.9 pN. The detailed deconvolution steps are shown for one run in the last row of Fig. 5. The intrinsic free-energy profile $\tilde{\mathcal{F}}_{\rm p}$ is shown for both runs in Fig. 5*H* (solid and dotted blue curves, respectively). Accounting for error due to finite trajectory length and uncertainties in the apparatus parameters, the median total uncertainty in each of the reconstructed landscapes is about 0.4 k_BT in the z range shown (see *SI Text* for full error analysis). The landscapes from the two independent runs have a median difference of 0.3 k_BT, and hence the method gives consistent results between runs, up to a small experimental uncertainty, an important test of its practical utility. The reproducibility of $\tilde{\mathcal{F}}_{\rm p}$ is a testament to the stability of the dual optical tweezer setup. Each trajectory lasted for more than 100 s, and thus collected ~(10^2–10^5) of the various types of transitions between protein states (the slowest transition, U → I2, occurred on time scales of 0.4–0.6 s).

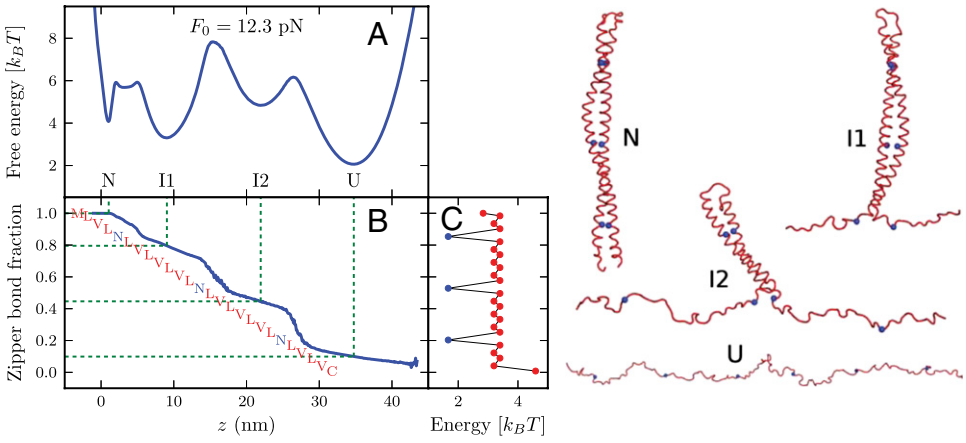

Fig. 4. Intrinsic characteristics of the LZ26 leucine zipper at constant F_0, derived from SOP simulations in the absence of handles/beads. (*A*) LZ26 free energy $\tilde{\mathcal{F}}_{\rm p}$ at $F_0 = 12.3$ pN vs. end-to-end extension z. (*Right*) Representative protein configurations from the four wells (N, I1, I2, U), with asparagine residues colored blue. (*B*) Average fraction of native contacts between the two alpha-helical strands of LZ26 (the "zipper bonds") as a function of z. (*Left*) Lists of the a and d residues in the heptads making up the amino acid sequence for each LZ26 strand, placed according to their position along the zipper. Asparagines (N) are highlighted in blue. (*C*) For the residues listed in *B*, the residue contact energies used in the SOP simulation [rescaled BT (30) values].

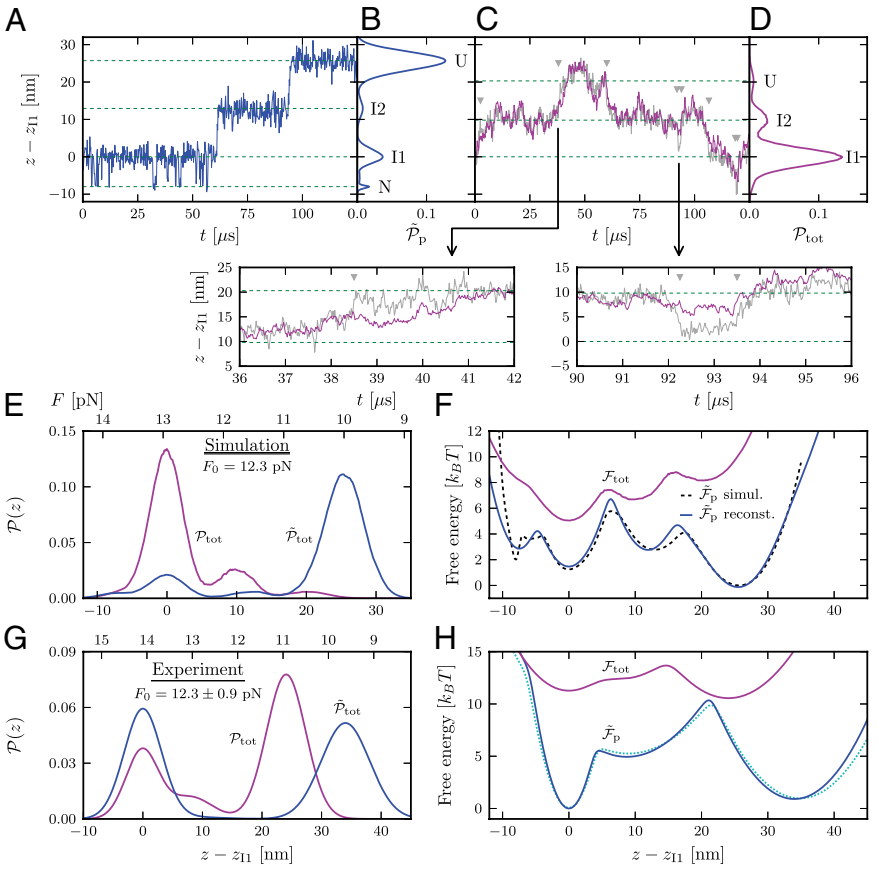

Fig. 5. (*A* and *B*) A trajectory fragment and the probability distribution $\tilde{\mathcal{P}}_p$ from SOP simulations of the LZ26 leucine zipper at constant force $F_0 = 12.3$ pN in the absence of handles/beads. (*C* and *D*) A trajectory fragment and the total system distribution $\mathcal{P}_{tot}$ at $z_{trap} = 503$ nm. *C* shows both the total extension $z_{tot}(t)$ (purple) and the protein extension $z_p(t)$ (gray). Triangles mark times when the protein makes a transition between states, and the arrows point to two enlarged portions of the trajectories. In all cases the z-axis origin is z_{I1}, the peak location of the I1 intermediate state. (*E*–*G*) Leucine zipper free-energy profiles extracted from time series (third row = simulation, fourth row = experiment). First column shows the total system end-to-end distribution $\mathcal{P}_{tot}$ and the corresponding $\tilde{\mathcal{P}}_{tot}$ at constant force $F_0 = 12.3$ pN. In the experimental case $F_0 = 12.3 \pm 0.9$ pN is the midpoint force at which the I1 and U states are equally likely. For $\mathcal{P}_{tot}$, $z_{trap} = 503$ nm (simulation), 1553 ± 1 nm (experiment). Force scales at the top are the range of trap forces for $\mathcal{P}_{tot}$. Second column shows the computed intrinsic protein free-energy profiles $\tilde{\mathcal{F}}_p$ compared with the total system profile, $\mathcal{F}_{tot}$ (shifted upward for clarity). (*F*) SOP simulations for the protein alone at constant F_0 provide a reference landscape, drawn as a dashed line. (*H*) Dotted curve is the reconstructed $\tilde{\mathcal{F}}_p$ at the midpoint force $F_0 = 12.1 \pm 0.9$ pN, from a second, independent experimental trajectory, with $z_{trap} = 1547 \pm 1$ nm. $\tilde{\mathcal{F}}_p$ curves have a median uncertainty of 0.4 k_BT over the plotted range (see *SI Text* for error analysis).

Comparison between the experimental $\tilde{\mathcal{F}}_p$ in Fig. 5*H* and the simulation result in Fig. 5*F* reveals a notable difference: The landscape constructed using the experimental data does not have four basins. The N state may not be discernible in the experiment because of the limited resolution of the apparatus (see below). The spacing between the I1 and I2 wells is similar in the simulation and experiment (~9–13 nm), but that between I2 and U is ~13 nm in the simulation vs. 25 nm in the experiment. This is likely due to a larger helix content in the unfolded state in the simulations.

Discussion

Origins of the Variance in the Point-Spread Function. Our theory for the point-spread function $\tilde{P}_{bh}$ can be used to understand the interplay of physical effects that relate the intrinsic protein distribution to the total system. To quantify the various contributions to $\tilde{P}_{bh}$, we calculated its variance. Because variances of probability distributions combine additively upon convolution, we break down the variance of $\tilde{P}_{bh}$ into the individual bead, handle, and linker contributions. Fig. 3*C* shows the fraction of the variance associated with each component as a function of the handle elastic modulus γ at $F_0 = 12.3$ pN, with $R_b = 500$ nm, $L = 188$ nm, and $l_p = 20$ nm (the approximate experimental parameters from ref. 11). For any given value of γ, the height of each of the four colored slices represents four fractions. Although not directly measured in ref. 11, we have assumed $\kappa = 200$ kcal/mol· nm^2, $\ell = 1.5$ nm for the linkers. The handle contribution is itself broken down into the "elastic" part, defined as the extra variance due to finite γ, compared with an inextensible ($\gamma \to \infty$) worm-like chain (WLC), and the remainder, which we call the WLC part. For the case of ref. 11, $\gamma = 400$ pN. Because the length extension relative to the WLC result is $\sim F_0/\gamma$, we expect finite handle extensibility to play a small role. Surprisingly, the elastic contribution to the total $\tilde{\mathcal{P}}_{bh}$ variance at this γ is 43%, comparable to the WLC contribution of 48%. Hence, in predicting $\tilde{\mathcal{P}}_{bh}$ correctly it is important to account for both the bending rigidity and elasticity of the handles.

Instrumental Noise Filtering and the Limits of the Theoretical Approach. The difference in the number of wells in the simulation and experimental free-energy landscapes of the leucine zipper is related to finite time and spatial resolution. The measured time series is subject to noise and low-pass filtering due to the apparatus (15). The standard experimental protocol often involves additional low-pass filtering as a way of removing noise and smoothing trajectories: for the leucine zipper every five data points (originally recorded at 10-μs intervals) are averaged together during collection to give a time step of 50 μs (11). Noise broadens the measured distribution of bead separations, whereas low-pass filtering narrows it. The finite bandwidth scaling (FBS) technique (*Materials and Methods*; *SI Text*), based on the details of the specific apparatus used in the experiment, estimates and corrects for the distortions.

The FBS theory can only apply corrections to peaks (i.e., distinct protein states) that we observe in the measured probability distributions. There is the possibility of protein states leaving no discernible signature in the recorded distribution. The N state in the leucine zipper is connected only to the I1 state in the folding pathway. In the simulations, where the N state is directly observed, it has short mean lifetimes ($\lesssim 6$ μs in the studied force range). The N ↔ I1 change involves the shortest mean extension difference (~8 nm) among all of the transitions. If the N state in the actual protein has similar properties, it could be impossible to resolve it in the experimental data for two reasons: (*i*) Regardless of any additional filtering, the intrinsic low-pass characteristics of the apparatus filter out states with very short lifetimes. For our LOT setup, the effective low-pass filter time scale for the detectors/electronics is $\tau_f \sim 7$ μs (*SI Text*), which is at the cutting edge of current technology. Thus, states

with lifetimes $\lesssim \tau_f$ will not appear as distinct peaks in the measured distribution. (*ii*) Independent of the filtering issues in detection/recording, environmental background noise in the time series also poses a problem, particularly because we measure bead displacements, and these have signal amplitudes at high frequencies that are generally attenuated compared with the intrinsic amplitudes of the protein conformational changes. The reason for this is that the beads have much larger hydrodynamic drag than dsDNA handles or proteins, and their characteristic relaxation times τ_r in the optical traps may be comparable to or larger than the lifetime of a particular protein state. The bead cannot fully respond to force changes on time scales shorter than its relaxation time (14). For example, $\tau_r \sim 20$ μs in the leucine zipper experiment. If the lifetime of the N state at a particular force is much smaller than τ_r, protein transitions from I1→N→I1 will generally occur before the bead can relax into the N-state equilibrium position. If the bead displacements associated with these transitions are smaller than the noise amplitude in the time series, the entire excursion to the N state will be lost to the noise.

We can illustrate the finite response time of the bead using simulations where resolution is not limited by noise or apparatus filtering, allowing us to see the relationship between $z_{\rm tot}(t)$ and $z_{\rm p}(t)$, compared in two different trajectory fragments in Fig. 5*C*. Triangles in the figure indicate times where the protein makes a transition between states. Changes in protein extension during these transitions are very rapid, and the bead generally mirrors these changes with a small time lag, as seen in the enlarged trajectory interval at $t = 36$–42 μs. When the protein makes sharp, extremely brief excursions (such as a visit to the N state from I1 in the enlarged interval $t = 90$–96 μs), the corresponding changes in bead separation are smaller and much less well-defined. In the presence of noise, such tiny changes would be obscured.

Thus, we surmise that the N state is not observable due to some combination of apparatus filtering, noise, and finite bead response time. Hence, the theory applied to the experimental data produces a landscape with only I1, I2, and U wells, as opposed to the four wells in the simulation data.

Conclusions

Extraction of the energy landscape of biomolecules using LOT data is complicated because accurate analysis depends on correcting for distortions due to system components. We have solved this problem completely by developing a theoretically based construction method that accounts for these factors. Through an array of tests involving an analytically solvable hairpin model, coarse-grained protein simulations, and experimental data, we have demonstrated the robustness of the technique in a range of realistic scenarios. The method works for arbitrarily complicated landscapes, producing consistent results when the same protein is studied under different force scales.

Materials and Methods

FBS. Probability distributions derived from experimental time series of bead–bead separations are corrupted by noise, finite apparatus bandwidth, and in some cases additional filtering due to the data processing protocol. We developed FBS theory to model and correct for these effects (*SI Text*), using information encoded in time series autocorrelations, together with spectral characterization of the dual-trap optical tweezer detector and electronic systems (15). All of the experimental distributions $\mathcal{P}_{\rm tot}$ in the main text were first processed by FBS.

Leucine Zipper. We use a variant of the coarse-grained SOP model (29, 31), where each of the 176 residues in LZ26 is represented by a bead centered at the C_α position (*SI Text*). The α-helical secondary structure is stabilized by interactions which mimic (i, $i + 4$) hydrogen bonding (32). We use residue-dependent energies for tertiary interactions (30). Simulations are carried out using an overdamped Brownian dynamics algorithm (33).

ACKNOWLEDGMENTS. M.H. was a Ruth L. Kirschstein National Research Service Postdoctoral Fellow supported by Grant 1 F32 GM 97756-1 from the National Institute of General Medical Sciences. D.T. was supported by Grant GM 089685 from the National Institutes of Health.

1. Thirumalai D, Hyeon C (2005) RNA and protein folding: Common themes and variations. *Biochemistry* 44(13):4957–4970.
2. Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: The energy landscape perspective. *Annu Rev Phys Chem* 48:545–600.
3. Dill KA, Ozkan SB, Shell MS, Weikl TR (2008) The protein folding problem. *Annu Rev Biophys* 37:289–316.
4. Thirumalai D, O'Brien EP, Morrison G, Hyeon C (2010) Theoretical perspectives on protein folding. *Annu Rev Biophys* 39:159–183.
5. Guo Z, Thirumalai D (1995) Kinetics of protein folding: Nucleation mechanism, time scales, and pathways. *Biopolymers* 36(1):83–102.
6. Klimov DK, Thirumalai D (2005) Symmetric connectivity of secondary structure elements enhances the diversity of folding pathways. *J Mol Biol* 353(5):1171–1186.
7. Stigler J, Ziegler F, Gieseke A, Gebhardt JCM, Rief M (2011) The complex folding network of single calmodulin molecules. *Science* 334(6055):512–516.
8. Woodside MT, et al. (2006) Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* 314(5801):1001–1004.
9. Woodside MT, et al. (2006) Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc Natl Acad Sci USA* 103(16):6190–6195.
10. Neupane K, Yu H, Foster DAN, Wang F, Woodside MT (2011) Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism. *Nucleic Acids Res* 39(17):7677–7687.
11. Gebhardt JCM, Bornschlögl T, Rief M (2010) Full distance-resolved folding energy landscape of one single protein molecule. *Proc Natl Acad Sci USA* 107(5):2013–2018.
12. Elms PJ, Chodera JD, Bustamante C, Marqusee S (2012) The molten globule state is unusually deformable under mechanical force. *Proc Natl Acad Sci USA* 109(10): 3796–3801.
13. Hyeon C, Thirumalai D (2006) Forced-unfolding and force-quench refolding of RNA hairpins. *Biophys J* 90(10):3410–3427.
14. Manosas M, et al. (2007) Force unfolding kinetics of RNA using optical tweezers. II. Modeling experiments. *Biophys J* 92(9):3010–3021.
15. von Hansen Y, Mehlich A, Pelz B, Rief M, Netz RR (2012) Auto- and cross-power spectral analysis of dual trap optical tweezer experiments using Bayesian inference. *Rev Sci Instrum* 83(9):095116.
16. Yu H, et al. (2012) Direct observation of multiple misfolding pathways in a single prion protein molecule. *Proc Natl Acad Sci USA* 109(14):5283–5288.
17. Hummer G, Szabo A (2010) Free energy profiles from single-molecule pulling experiments. *Proc Natl Acad Sci USA* 107(50):21441–21446.
18. Hinczewski M, von Hansen Y, Netz RR (2010) Deconvolution of dynamic mechanical networks. *Proc Natl Acad Sci USA* 107(50):21493–21498.
19. Hyeon C, Morrison G, Thirumalai D (2008) Force-dependent hopping rates of RNA hairpins can be estimated from accurate measurement of the folding landscapes. *Proc Natl Acad Sci USA* 105(28):9604–9609.
20. Greenleaf WJ, Woodside MT, Abbondanzieri EA, Block SM (2005) Passive all-optical force clamp for high-resolution laser trapping. *Phys Rev Lett* 95(20):208102.
21. Neuman KC, Block SM (2004) Optical trapping. *Rev Sci Instrum* 75(9):2787–2809.
22. Ferrenberg AM, Swendsen RH (1989) Optimized Monte Carlo data analysis. *Phys Rev Lett* 63(12):1195–1198.
23. Shirts MR, Chodera JD (2008) Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys* 129(12):124105.
24. de Messieres M, Brawn-Cinani B, La Porta A (2011) Measuring the folding landscape of a harmonically constrained biopolymer. *Biophys J* 100(11):2736–2744.
25. Kierfeld J, Niamploy O, Sa-yakanit V, Lipowsky R (2004) Stretching of semiflexible polymers with elastic bonds. *Eur Phys J E Soft Matter* 14(1):17–34.
26. Bornschlögl T, Rief M (2006) Single molecule unzipping of coiled coils: Sequence resolved stability profiles. *Phys Rev Lett* 96(11):118102.
27. O'Shea EK, Klemm JD, Kim PS, Alber T (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* 254(5031):539–544.
28. Bornschlögl T, Rief M (2008) Single-molecule dynamics of mechanical coiled-coil unzipping. *Langmuir* 24(4):1338–1342.
29. Hyeon C, Dima RI, Thirumalai D (2006) Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins. *Structure* 14(11):1633–1645.
30. Betancourt MR, Thirumalai D (1999) Pair potentials for protein folding: Choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci* 8(2):361–369.
31. Mickler M, et al. (2007) Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations. *Proc Natl Acad Sci USA* 104(51): 20268–20273.
32. Denesyuk NA, Thirumalai D (2011) Crowding promotes the switch from hairpin to pseudoknot conformation in human telomerase RNA. *J Am Chem Soc* 133(31): 11858–11861.
33. Ermak DL, McCammon JA (1978) Brownian dynamics with hydrodynamic interactions. *J Chem Phys* 69(4):1352–1360.

Supporting Information

Hinczewski et al. 10.1073/pnas.1214051110

SI Text

1. Theory for Free-Energy Construction from Mechanical Folding Time Series

1.1. Optical Trap Hamiltonian. We begin with the Hamiltonian for the beads in the traps (Fig. 1), which allows us to introduce the relevant variables of the system. If the displacements of the beads from the trap centers are small (<100 nm for a laser of 1,064-nm wavelength and bead radii $\sim\mathcal{O}$(100 nm) (1), the trap Hamiltonian can be approximately written as

$$\mathcal{H}_{\rm trap}(\mathbf{r}_1,\mathbf{r}_2)=\frac{1}{2}k_x\left(x_1^2+x_2^2\right)+\frac{1}{2}k_y\left(y_1^2+y_2^2\right)+\frac{1}{2}k_z\left[z_1^2+\left(z_{\rm trap}-z_2\right)^2\right], \quad [S1]$$

where $\mathbf{r}_i=(x_i,y_i,z_i)$ is the position of the ith bead center, k_x, k_y, k_z are the trap strengths along each coordinate direction, and the two traps are positioned at $z=0$ and $z=z_{\rm trap}$, respectively. Given the cylindrical symmetry of the optical traps around the $\hat{\mathbf{y}}$ axis, we take $k_x=k_z\equiv k_{\rm trap}$ and $k_y=\alpha k_{\rm trap}$, where the weaker axial trapping is reduced by a factor $\alpha<1$ (2). We assume both traps have equal strengths, although our method can be generalized to an asymmetric configuration, where the two traps have different strengths $k_{{\rm trap},1}\neq k_{{\rm trap},2}$. In this case the reconstruction procedure derived below is valid with the substitution $k_{\rm trap}=2k_{{\rm trap},1}k_{{\rm trap},2}/(k_{{\rm trap},1}+k_{{\rm trap},2})$.

We rewrite the Hamiltonian in Eq. **S1** by defining a total end-to-end coordinate $\mathbf{r}_{\rm tot}\equiv\mathbf{r}_2-\mathbf{r}_1=(x_{\rm tot},y_{\rm tot},z_{\rm tot})$, and a total center-of-mass coordinate $\mathbf{R}_{\rm tot}\equiv\mathbf{r}_2+\mathbf{r}_1=(X_{\rm tot},Y_{\rm tot},Z_{\rm tot})$. In terms of these variables, $\mathcal{H}_{\rm trap}$ becomes

$$\begin{aligned}\mathcal{H}_{\rm trap}(\mathbf{r}_1,\mathbf{r}_2)&=\mathcal{H}^{\rm cm}_{\rm trap}(\mathbf{R}_{\rm tot})+\mathcal{H}^{\rm ee}_{\rm trap}(\mathbf{r}_{\rm tot}),\\ \mathcal{H}^{\rm cm}_{\rm trap}(\mathbf{R}_{\rm tot})&=\frac{1}{4}k_xX_{\rm tot}^2+\frac{1}{4}k_yY_{\rm tot}^2+\frac{1}{4}k_z\left(z_{\rm trap}-Z_{\rm tot}\right)^2,\\ \mathcal{H}^{\rm ee}_{\rm trap}(\mathbf{r}_{\rm tot})&=\frac{1}{4}k_xx_{\rm tot}^2+\frac{1}{4}k_yy_{\rm tot}^2+\frac{1}{4}k_z\left(z_{\rm trap}-z_{\rm tot}\right)^2.\end{aligned} \quad [S2]$$

The variables $z_{\rm tot}$ and $z_{\rm trap}$ are explicitly labeled in Fig. 1.

1.2. Equilibrium Distribution of the System. The equilibrium probability $\mathcal{P}_{\rm tot}(\mathbf{R}_{\rm tot},\mathbf{r}_{\rm tot})$ of finding the beads at positions with a given $\mathbf{R}_{\rm tot}$ and $\mathbf{r}_{\rm tot}$ can be expressed as

$$\mathcal{P}_{\rm tot}(\mathbf{R}_{\rm tot},\mathbf{r}_{\rm tot})=Ae^{-\beta\mathcal{H}^{\rm cm}_{\rm trap}(\mathbf{R}_{\rm tot})-\beta\mathcal{H}^{\rm ee}_{\rm trap}(\mathbf{r}_{\rm tot})}\mathcal{Q}_{\rm tot}(\mathbf{r}_{\rm tot}), \quad [S3]$$

where $\beta=1/k_BT$, A is a normalization constant, and $\mathcal{Q}_{\rm tot}(\mathbf{r}_{\rm tot})$ is the equilibrium probability of the total bead–handle–protein system having bead separation $\mathbf{r}_{\rm tot}$ in the absence of the external trapping potential or any applied force. By translational symmetry $\mathcal{Q}_{\rm tot}$ is independent of the center-of-mass coordinates, and by rotational symmetry $\mathcal{Q}_{\rm tot}(\mathbf{r}_{\rm tot})=\mathcal{Q}_{\rm tot}(|\mathbf{r}_{\rm tot}|)$. Thus, if we introduce cylindrical coordinates $\mathbf{r}_{\rm tot}=(\rho_{\rm tot},\phi_{\rm tot},z_{\rm tot})$, where $\rho_{\rm tot}=\sqrt{x_{\rm tot}^2+y_{\rm tot}^2}$, $\phi_{\rm tot}=\tan^{-1}(y_{\rm tot}/x_{\rm tot})$, there is no angular dependence, so that $\mathcal{Q}_{\rm tot}(\mathbf{r}_{\rm tot})=\mathcal{Q}_{\rm tot}(\rho_{\rm tot},z_{\rm tot})$. We are ultimately interested in the marginal probability $\mathcal{P}_{\rm tot}(z_{\rm tot})$, which can be derived from the experimental time series and forms the starting point of our theoretical procedure to obtain the desired free-energy profile. We obtain $\mathcal{P}_{\rm tot}(z_{\rm tot})$ from $\mathcal{P}_{\rm tot}(\mathbf{R}_{\rm tot},\mathbf{r}_{\rm tot})$ by integrating over the $\mathbf{R}_{\rm tot}$, $\rho_{\rm tot}$, and $\phi_{\rm tot}$ degrees of freedom:

$$\begin{aligned}\mathcal{P}_{\rm tot}(z_{\rm tot})&\equiv\int\rho_{\rm tot}d\rho_{\rm tot}d\phi_{\rm tot}d\mathbf{R}_{\rm tot}\mathcal{P}_{\rm tot}(\mathbf{R}_{\rm tot},\mathbf{r}_{\rm tot})\\ &=B\int\rho_{\rm tot}d\rho_{\rm tot}d\phi_{\rm tot}e^{-\beta\mathcal{H}^{\rm ee}_{\rm trap}(\rho_{\rm tot},\phi_{\rm tot},z_{\rm tot})}\mathcal{Q}_{\rm tot}(\rho_{\rm tot},z_{\rm tot})\\ &=2\pi B\int\rho_{\rm tot}d\rho_{\rm tot}e^{-\frac{1}{8}\beta\left(k_x+k_y\right)\rho_{\rm tot}^2-\frac{1}{4}\beta k_z\left(z_{\rm trap}-z_{\rm tot}\right)^2}\\ &\quad\times I_0\left(\frac{1}{8}\left(k_x-k_y\right)\rho_{\rm tot}^2\right)\mathcal{Q}_{\rm tot}(\rho_{\rm tot},z_{\rm tot})\\ &\approx Ce^{-\frac{1}{4}\beta k_z\left(z_{\rm trap}-z_{\rm tot}\right)^2}\mathcal{Q}_{\rm tot}(z_{\rm tot}).\end{aligned} \quad [S4]$$

Here, B and C are constants that have absorbed the result of integrating over $\mathbf{R}_{\rm tot}$ and $\rho_{\rm tot}$, respectively, and I_0 is a modified Bessel function of the first kind. Up to the third line the calculation in Eq. **S4** is exact. In the last step we make the problem fully one-dimensional by approximately relating $\mathcal{P}_{\rm tot}(z_{\rm tot})$ to $\mathcal{Q}_{\rm tot}(z_{\rm tot})$, defined as $\mathcal{Q}_{\rm tot}(z_{\rm tot})=\int\rho_{\rm tot}d\rho_{\rm tot}\mathcal{Q}_{\rm tot}(\rho_{\rm tot},z_{\rm tot})$. We are forced to make this crucial approximation, because experiments have access only to the $\hat{\mathbf{z}}$ fluctuations through $\mathcal{P}_{\rm tot}(z_{\rm tot})$, but generally do not have complete information about the transverse components. As mentioned in the main text, the last step in Eq. **S4** becomes exact if: (*i*) $k_x=k_y=0$; or (*ii*) when $\mathcal{Q}_{\rm tot}(\rho_{\rm tot},z_{\rm tot})$ is separable in the form $\mathcal{Q}_{\rm tot}(\rho_{\rm tot},z_{\rm tot})=f(\rho_{\rm tot})\mathcal{Q}_{\rm tot}(z_{\rm tot})$ for some function f. Although condition (*ii*) is not expected to be generally valid, we can approximately satisfy (*i*) when $\beta k_{\rm trap}\overline{\rho}_{\rm tot}^2\ll1$, where $\overline{\rho}_{\rm tot}$ is the typical length scale of total system fluctuations transverse to $\hat{\mathbf{z}}$. Thus, for sufficiently soft traps, we have in Eq. **S4** a useful relation between the $\hat{\mathbf{z}}$ marginal probabilities of the total system with and without the external trapping potentials.

1.3. Convolution. Because $\mathcal{Q}_{\rm tot}(z_{\rm tot})$ is the total end-to-end z-component distribution in the absence of any external trapping potential or applied force, the corresponding distribution for the total system with constant tension F_0 applied to the beads along $\hat{\mathbf{z}}$ is given by $\tilde{\mathcal{P}}_{\rm tot}(z_{\rm tot};F_0)=\exp(\beta F_0z_{\rm tot})\mathcal{Q}_{\rm tot}(z_{\rm tot})$. Substituting for $\mathcal{Q}_{\rm tot}(z_{\rm tot})$ using Eq. **S4**, we find the following relation for $\tilde{\mathcal{P}}_{\rm tot}(z_{\rm tot};F_0)$, which constitutes step i of our construction procedure in the main text:

$$\tilde{\mathcal{P}}_{\rm tot}(z_{\rm tot};F_0)\approx C^{-1}e^{\beta F_0z+\frac{1}{4}\beta k_z\left(z_{\rm trap}-z_{\rm tot}\right)^2}\mathcal{P}_{\rm tot}(z_{\rm tot}). \quad [S5]$$

The quantity $\mathcal{P}_{\rm tot}(z_{\rm tot})$ on the right-hand side can be derived from the experimental time series, and thus Eq. **S5** allows us to obtain an equilibrium distribution in the constant-force ensemble $\tilde{\mathcal{P}}_{\rm tot}(z_{\rm tot};F_0)$ directly from the folding trajectories.

In the constant-force ensemble, $\tilde{\mathcal{P}}_{\rm tot}$ is just a 1D convolution of the probabilities of the individual system components:

$$\tilde{\mathcal{P}}_{\rm tot}=\tilde{\mathcal{P}}_{\rm b}*\tilde{\mathcal{P}}_{\rm h}*\tilde{\mathcal{P}}_{\rm p}*\tilde{\mathcal{P}}_{\rm h}*\tilde{\mathcal{P}}_{\rm b}, \quad [S6]$$

where "*" denotes the convolution operator. The probability $\tilde{\mathcal{P}}_\lambda(z_\lambda;F_0)$ is the equilibrium distribution of z_λ at constant force F_0, where λ denotes a bead, handle, or protein. The quantity z_λ is the end-to-end distance of λ along $\hat{\mathbf{z}}$. Using the notation in Fig. 1, we can give a few examples: for the protein $z_{\rm p}=(\mathbf{r}_{p2}-\mathbf{r}_{p1})\cdot\hat{\mathbf{z}}$; for the left handle $z_{\rm h}=(\mathbf{r}_{p1}-\mathbf{r}_1')\cdot\hat{\mathbf{z}}$; for the left bead $z_{\rm b}=(\mathbf{r}_1'-\mathbf{r}_1)\cdot\hat{\mathbf{z}}$. In Fourier space, Eq. **S6**, which is the key equation for step *ii* of the construction procedure in the main text, has a simple form:

$$\tilde{\mathcal{P}}_{\rm tot}(k;F_0)=\tilde{\mathcal{P}}_{\rm b}^2(k;F_0)\tilde{\mathcal{P}}_{\rm h}^2(k;F_0)\tilde{\mathcal{P}}_{\rm p}(k;F_0)\equiv\tilde{\mathcal{P}}_{\rm bh}(k;F_0)\tilde{\mathcal{P}}_{\rm p}(k;F_0), \quad \textbf{[S7]}$$

where $\tilde{\mathcal{P}}_\lambda(k;F_0)$ is the Fourier transform of $\tilde{\mathcal{P}}_\lambda(z_\lambda;F_0)$. Here, $\tilde{\mathcal{P}}_{\rm bh}(k;F_0)=\tilde{\mathcal{P}}_{\rm b}^2(k;F_0)\tilde{\mathcal{P}}_{\rm h}^2(k;F_0)$ is the Fourier transform of the convolution of all of the bead and handle distributions. If the left and right handles (or analogously the beads) had distinct properties (i.e., different sizes), then the factor $\tilde{\mathcal{P}}_{\rm h}^2(k;F_0)$ in $\tilde{\mathcal{P}}_{\rm bh}(k;F_0)$ would be replaced by the product $\tilde{\mathcal{P}}_{\rm h1}(k;F_0)\tilde{\mathcal{P}}_{\rm h2}(k;F_0)$ of the distinct handle terms. Given the rotational properties of the beads and modeling the handles as semiflexible polymers, we can derive a numerically exact form for the Fourier components $\tilde{\mathcal{P}}_{\rm bh}(k;F_0)$, and hence by inversion the corresponding real-space distribution. This will allow us to directly recover $\tilde{\mathcal{P}}_{\rm p}$ from $\tilde{\mathcal{P}}_{\rm tot}$ without resorting to an experimental estimate for the point-spread function, which is problematic due to the varying force conditions that arise in optical traps with nonzero stiffness.

1.4. Bead Distribution. The first step in finding $\tilde{\mathcal{P}}_{\rm bh}(k;F_0)=\tilde{\mathcal{P}}_{\rm b}^2(k;F_0)\tilde{\mathcal{P}}_{\rm h}^2(k;F_0)$ is to obtain an expression for the Fourier-space bead probability $\tilde{\mathcal{P}}_{\rm b}(k;F_0)$. Taking as an example the left bead in Fig. 1, let $\mathbf{r}_{\rm b}=\mathbf{r}_1'-\mathbf{r}_1$ be the vector between the bead center and the point on the bead surface that is attached to the handle. This vector has a fixed length R_b given by the bead radius, but its direction can fluctuate, subject to a constant force F_0 along $\hat{\mathbf{z}}$. The equilibrium distribution $\tilde{\mathcal{P}}_{\rm b}(\mathbf{r}_{\rm b};F_0)$ is given by

$$\tilde{\mathcal{P}}_{\rm b}(\mathbf{r}_{\rm b};F_0)=A_{\rm b}e^{\beta F_0 z_{\rm b}}\delta(|\mathbf{r}_{\rm b}|-R_b), \quad \textbf{[S8]}$$

with the delta function enforcing the constraint $|\mathbf{r}_{\rm b}|=R_b$, and the normalization constant A_b. The quantity $\tilde{\mathcal{P}}_{\rm b}(k;F_0)$ is the Fourier transform of $\tilde{\mathcal{P}}_{\rm b}(\mathbf{r}_{\rm b};F_0)$ evaluated at $\mathbf{k}=k\hat{\mathbf{z}}$:

$$\begin{aligned}\tilde{\mathcal{P}}_{\rm b}(k;F_0)&=\int d\mathbf{r}_{\rm b}e^{-ikz_{\rm b}}\tilde{\mathcal{P}}_{\rm b}(\mathbf{r}_{\rm b};F_0)\\&=\frac{\beta F_0\sinh((\beta F_0-ik)R_b)}{(\beta F_0-ik)\sinh(\beta F_0R_b)}.\end{aligned} \quad \textbf{[S9]}$$

1.5. Handle Distribution. Although the Fourier components $\tilde{\mathcal{P}}_{\rm h}(k;F_0)$ of the semiflexible handle distribution do not have a simple analytic expression, they can be calculated numerically to arbitrary accuracy. The Hamiltonian for the semiflexible handle polymer with contour length L, persistence length l_p, and elastic stretching modulus γ can be exactly mapped onto the propagator of a quantum particle on the surface of a unit sphere (3, 4). Following the approach in ref. 4, we describe the polymer as a continuous spatial contour $\mathbf{r}(s)$ in terms of an unstretched arc length l_s, which runs from $s=0$ to $s=L$. At each point s we define a unit tangent vector $\mathbf{u}(s)$. The end-to-end distance $\mathbf{r}(L)-\mathbf{r}(0)$ can be written as

$$\mathbf{r}(L)-\mathbf{r}(0)=\int_0^L ds(1+\epsilon(s))\mathbf{u}(s), \quad \textbf{[S10]}$$

where $1+\epsilon(s)$ is the local relative bond length extension. For an inextensible ($\gamma\to\infty$) worm-like chain, $\epsilon(s)=0$ for all s, which corresponds to all bonds in the chain having fixed length. For finite γ, the $\epsilon(s)$ are additional degrees of freedom in the system, which together with the unit tangent vectors $\mathbf{u}(s)$ completely define the contour. The Hamiltonian $\mathcal{H}(\mathbf{u}(s),\epsilon(s))$ for the semiflexible polymer under tension is

$$\begin{aligned}\beta\mathcal{H}(\mathbf{u}(s),\epsilon(s))&=\frac{l_p}{2}\int_0^L ds(\partial_s\mathbf{u}(s))^2-f\hat{\mathbf{z}}\cdot(\mathbf{r}(L)-\mathbf{r}(0))+\frac{\beta\gamma}{2}\int_0^L ds\epsilon^2(s),\\&=\int_0^L ds\left[\frac{l_p}{2}(\partial_s\mathbf{u}(s))^2-f(1+\epsilon(s))u_z(s)+\frac{\beta\gamma}{2}\epsilon^2(s)\right],\end{aligned} \quad \textbf{[S11]}$$

where $\beta=1/k_BT$ and we have used Eq. **S10** for the end-to-end distance. The first term in Eq. **S11** corresponds to a bending energy parametrized by the persistence length l_p, the second term is due to an applied mechanical force k_BTf along $\hat{\mathbf{z}}$, and the third term describes the stretching energy of the bonds, with elastic modulus γ. For prestretching tension F_0, $f=\beta F_0$, but for convenience we will extend the definition of $\mathcal{H}$ to include arbitrary f to obtain the Fourier components of the end-to-end probability distribution below.

The partition function of the polymer (with free end boundary conditions) can be expressed as a path integral over all possible configurations of $\mathbf{u}(s)$ and $\epsilon(s)$, with the constraint that $\mathbf{u}^2(s)=1$ at each s:

$$\begin{aligned}Z_{\rm h}(f)&=\int\mathcal{D}\mathbf{u}(s)\prod_s\delta(\mathbf{u}^2(s)-1)\int\mathcal{D}\epsilon(s)\exp[-\beta\mathcal{H}(\mathbf{u}(s),\epsilon(s))],\\&\equiv\int\mathcal{D}\mathbf{u}(s)\prod_s\delta(\mathbf{u}^2(s)-1)\exp\left[-\beta\mathcal{H}_{\rm eff}(\mathbf{u}(s))\right],\end{aligned} \quad \textbf{[S12]}$$

up to some normalization constant. In the second line we have carried out the path integral over $\epsilon(s)$ exactly to express $Z_{\rm h}(f)$ in terms of an effective Hamiltonian $\mathcal{H}_{\rm eff}(\mathbf{u}(s))$ depending on the tangent vectors alone,

$$\beta\mathcal{H}_{\rm eff}(\mathbf{u}(s))=\int_0^L ds\left[\frac{l_p}{2}(\partial_s\mathbf{u}(s))^2-fu_z(s)-\frac{f^2}{2\beta\gamma}u_z^2(s)\right]. \quad \textbf{[S13]}$$

The probability of finding the polymer in a configuration with an end-to-end extension $z_{\rm h}$ along $\hat{\mathbf{z}}$ is given by (3)

$$\begin{aligned}\tilde{\mathcal{P}}_{\rm h}(z_{\rm h};F_0)&=\frac{1}{Z_{\rm h}(\beta F_0)}\int\mathcal{D}\mathbf{u}(s)\prod_s\delta(\mathbf{u}^2(s)-1)\\&\int\mathcal{D}\epsilon(s)\delta\left(z_{\rm hand}-\int_0^L ds(1+\epsilon(s))u_z(s)\right)\exp[-\beta\mathcal{H}(\mathbf{u}(s),\epsilon(s))]\\&=\frac{1}{Z_{\rm h}(\beta F_0)}\int\mathcal{D}\mathbf{u}(s)\prod_s\delta(\mathbf{u}^2(s)-1)\\&\int\mathcal{D}\epsilon(s)\int\frac{dk}{2\pi}e^{ik\left(z_{\rm h}-\int_0^L ds(1+\epsilon(s))u_z(s)\right)}\exp[-\beta\mathcal{H}(\mathbf{u}(s),\epsilon(s))]\\&\equiv\int\frac{dk}{2\pi}e^{ikz_{\rm h}}\tilde{\mathcal{P}}_{\rm h}(k;F_0),\end{aligned} \quad \textbf{[S14]}$$

where the Fourier components of the probability distribution are

$$\tilde{\mathcal{P}}_{\rm h}(k;F_0)=\frac{Z_{\rm h}(\beta F_0-ik)}{Z_{\rm h}(\beta F_0)}. \quad \textbf{[S15]}$$

To evaluate $\tilde{\mathcal{P}}_{\rm h}(k;F_0)$, we need to calculate $Z_{\rm h}(f)$. Let us define the propagator $G(\mathbf{u}_0, \mathbf{u}_L; L)$ as the path integral over all configurations with initial tangent $\mathbf{u}(0) = \mathbf{u}_0$ and final tangent $\mathbf{u}(L) = \mathbf{u}_L$:

$$G(\mathbf{u}_0,\mathbf{u}_L;L)=\int_{\mathbf{u}(0)=\mathbf{u}_0}^{\mathbf{u}(L)=\mathbf{u}_L}\mathcal{D}\mathbf{u}(s)\prod_s\delta\big(\mathbf{u}^2(s)-1\big)e^{-\beta\mathcal{H}_{\rm eff}(\mathbf{u}(s))}. \quad \textbf{[S16]}$$

This is related to the partition function through $Z_{\rm h}(f) = (4\pi)^{-2}\int_S d\mathbf{u}_0 d\mathbf{u}_L G(\mathbf{u}_0, \mathbf{u}_L; L)$, where the integrations are over the unit sphere S.

The quantum Hamiltonian corresponding to $\beta\mathcal{H}_{\rm eff}$ is

$$\mathcal{H}^{\rm qu}_{\rm eff}(f)=-\frac{1}{2l_p}\nabla^2-f\cos\theta-\frac{f^2}{2\beta\gamma}\cos^2\theta, \quad \textbf{[S17]}$$

describing a particle on the surface of a unit sphere, with $\theta = 0$ defining the $\hat{\mathbf{z}}$ direction. The propagator G can be written in terms of the quantum eigenvalues E_n and eigenstates $\psi_n(\mathbf{u})$ of $\mathcal{H}^{\rm qu}_{\rm eff}$:

$$\begin{aligned}G(\mathbf{u}_0,\mathbf{u}_L;L)&=\sum_n e^{-E_nL}\psi_n^*(\mathbf{u}_0)\psi_n(\mathbf{u}_L)\\&=\sum_{n,l,m,l',m'}e^{-E_nL}a^*_{nl'm'}a_{nlm}Y^*_{l'm'}(\mathbf{u}_0)Y_{lm}(\mathbf{u}_L),\end{aligned} \quad \textbf{[S18]}$$

where we have expanded the eigenstates in the basis of spherical harmonics, $\psi_n(\mathbf{u}) = \sum_{lm} a_{nlm} Y_{lm}(\mathbf{u})$. The coefficients a_{nlm} are the components of the nth eigenvector of the Hamiltonian $\mathcal{H}^{\rm qu}_{\rm eff}$ in the Y_{lm} basis. The partition function $Z_{\rm h}(f)$ becomes

$$\begin{aligned}Z_{\rm h}(f)&=\frac{1}{(4\pi)^2}\int_S d\mathbf{u}_0\,d\mathbf{u}_L\sum_{n,l,m,l',m'}e^{-E_nL}a^*_{nl'm}a_{nlm}Y^*_{l'm'}(\mathbf{u}_0)Y_{lm}(\mathbf{u}_L)\\&=\frac{1}{4\pi}\sum_n e^{-E_nL}a^*_{n00}a_{n00}\\&=\langle 0|e^{-L\mathcal{H}^{\rm qu}_{\rm eff}(f)}|0\rangle.\end{aligned} \quad \textbf{[S19]}$$

In the last step we have written the expression as a single component of the exponentiated matrix $\mathcal{H}^{\rm qu}_{\rm eff}(f)$ in the (l, m) spherical harmonic basis, where $|l\rangle$ denotes a state $(l, 0)$. Because the Hamiltonian matrix in the $m = 0$ subspace does not couple to $m \neq 0$ components, we only need $m = 0$ matrix elements to evaluate $Z_{\rm h}(f)$. The list of nonzero matrix entries in the $m = 0$ subspace is

$$\begin{aligned}\langle l|\mathcal{H}^{\rm qu}_{\rm eff}(f)|l\rangle&=\frac{1}{2l_p}l(l+1)-\frac{f^2}{2\beta\gamma}\frac{2l^2+2l-1}{(2l-1)(2l+3)},\\&\qquad l=0,1,2,\ldots,\\\langle l|\mathcal{H}^{\rm qu}_{\rm eff}(f)|l+1\rangle&=\langle l+1|\mathcal{H}^{\rm qu}_{\rm eff}(f)|l\rangle=\frac{f(l+1)}{\sqrt{(2l+1)(2l+3)}},\\&\qquad l=0,1,2,\ldots,\\\langle l|\mathcal{H}^{\rm qu}_{\rm eff}(f)|l+2\rangle&=\langle l+2|\mathcal{H}^{\rm qu}_{\rm eff}(f)|l\rangle=-\frac{f^2}{2\beta\gamma}\frac{(l+1)(l+2)}{(2l+3)\sqrt{(2l+1)(2l+5)}},\\&\qquad l=0,1,2,\ldots.\end{aligned} \quad \textbf{[S20]}$$

To carry out the matrix exponent, we truncate the matrix at $l_{\max} = 20$, which is sufficiently large for numerical accuracy.

In some experimental setups, covalent linkers are attached on both ends of each handle, connecting the handle to the neighboring bead and protein, as schematically drawn in Fig. S1*A*. The effect of linkers can be absorbed into the theory by modifying $\tilde{P}_{\rm h}(k;F_0)$. The simplest representation of a linker is a harmonic spring with stiffness κ and natural length ℓ. With one of these added at each end of the handle, Eq. **S15** becomes

$$\tilde{\mathcal{P}}_{\rm h}(k;F_0)=\frac{Z_{\rm h}(\beta F_0-ik)}{Z_{\rm h}(\beta F_0)}\frac{Z^2_{\rm link}(\beta F_0-ik;\kappa,\ell)}{Z^2_{\rm link}(\beta F_0;\kappa,\ell)}, \quad \textbf{[S21]}$$

where

$$\begin{aligned}Z_{\rm link}(f;\kappa,\ell)=\frac{1}{f}\sqrt{\frac{\pi}{2(\beta\kappa)^3}}e^{\frac{f(f-2\beta\kappa)}{2\beta\kappa}}\Bigg[&\ell\beta\kappa\,\mathrm{erf}\left(\frac{f-\ell\beta\kappa}{\sqrt{2\beta\kappa}}\right)\\&+e^{2f\ell}(f+\ell\beta\kappa)\left(\mathrm{erf}\left(\frac{f+\ell\beta\kappa}{\sqrt{2\beta\kappa}}\right)+1\right)+f\,\mathrm{erfc}\left(\frac{f-\ell\beta\kappa}{\sqrt{2\beta\kappa}}\right)-\ell\beta\kappa\Bigg].\end{aligned} \quad \textbf{[S22]}$$

1.6. Numerical Deconvolution to Extract the Protein Distribution. The expressions given by Eqs. **S9** and **S21** completely determine the Fourier-transformed point-spread function $\tilde{\mathcal{P}}_{\rm bh}(k;F_0)$ at all k. Naively, one could use Eq. **S7** to write

$$\tilde{\mathcal{P}}_{\rm p}(k;F_0)=\frac{\tilde{\mathcal{P}}_{\rm tot}(k;F_0)}{\tilde{\mathcal{P}}_{\rm bh}(k;F_0)}. \quad \textbf{[S23]}$$

Because $\tilde{\mathcal{P}}_{\rm tot}(k;F_0)$ is derivable from the experimental times series, this would immediately yield $\tilde{\mathcal{P}}_{\rm p}(k;F_0)$, and after inversion the ultimate goal, $\tilde{\mathcal{P}}_{\rm p}(z_{\rm p};F_0)$. However, this direct deconvolution in Fourier space is numerically unstable (5) due to the effects of round-off noise and the denominator in the equation for $\tilde{\mathcal{P}}_{\rm p}(k;F_0)$ approaching zero at large k.

To work around this problem, we implement the deconvolution in real space by solving the following integral equation for $\tilde{\mathcal{P}}_{\rm p}$ (the real space version of Eq. **S7**):

$$\int dz_{\rm p}\tilde{\mathcal{P}}_{\rm bh}\big(z_{\rm tot}-z_{\rm p};F_0\big)\tilde{\mathcal{P}}_{\rm p}\big(z_{\rm p};F_0\big)=\tilde{\mathcal{P}}_{\rm tot}(z_{\rm tot};F_0). \quad \textbf{[S24]}$$

One way to approach Eq. **S24** is to approximate the integral as a matrix–vector product by discretizing the $z_{\rm tot}$ and $z_{\rm p}$ ranges. However, the convolution matrix corresponding to $\tilde{\mathcal{P}}_{\rm bh}$ is generally ill-conditioned, so direct inversion to find a solution is unfeasible. Alternatively, to obtain robust, smooth results for the deconvolution, we can rewrite Eq. **S24** by representing the three quantities $\tilde{\mathcal{P}}_{\rm p}$, $\tilde{\mathcal{P}}_{\rm bh}$, and $\tilde{\mathcal{P}}_{\rm tot}$ in terms of suitable fitting functions. Because these are all probability distributions, in practice we can approximate them to arbitrary precision as sums of Gaussians $g(z;\zeta,v) = (2\pi v)^{-1/2}\exp(-(z-\zeta)^2/(2v))$,

$$\tilde{\mathcal{P}}_\alpha(z_\alpha;F_0)=\sum_{i=1}^{N_\alpha}a_i^\alpha g\big(z_\alpha;\zeta_i^\alpha,v_i^\alpha\big), \quad \textbf{[S25]}$$

where α = p, bh, or tot. The number of Gaussians needed for each distribution N_α is chosen depending on the problem. The two sets of parameters $\{a_i^{\rm tot},\xi_i^{\rm tot},v_i^{\rm tot}\}$ and $\{a_i^{\rm bh},\xi_i^{\rm bh},v_i^{\rm bh}\}$ (which implicitly depend on F_0) are computed by fitting to the known functions $\tilde{\mathcal{P}}_{\rm tot}$ and $\tilde{\mathcal{P}}_{\rm bh}$. The goal of the procedure is then to use Eq. **S24** to solve for the parameter set $\{a_i^{\rm p},\xi_i^{\rm p},v_i^{\rm p}\}$ describing the unknown function $\tilde{\mathcal{P}}_{\rm p}$. For the cases discussed in the main text, choosing $N_\alpha = 2$–3 was sufficient to find solutions $\tilde{\mathcal{P}}_{\rm p}$ such that

the left- and right-hand sides of Eq. **S24** had a median deviation ≲1% over all z_{tot} where $\mathcal{P}_{tot}(z_{tot}) > 10^{-6}$.

The details of the solution procedure are as follows: we choose $N_p = N_{tot}$, so that Eq. **S24** can be approximated as a one-to-one convolution mapping each Gaussian in $\tilde{\mathcal{P}}_p$ into a corresponding Gaussian in $\tilde{\mathcal{P}}_{tot}$. For all $i = 1,\ldots,N_{tot}$, Eq. **S24** describes the following relationships between the amplitudes, positions, and variances of the Gaussians:

$$\begin{aligned} a_i^{tot} &\approx a_i^p, \\ \xi_i^{tot} &\approx \sum_{j=1}^{N_{bh}} a_j^{bh}\left(\xi_j^{bh} + \xi_i^p\right), \\ v_i^{tot} &\approx \sum_{j=1}^{N_{bh}} a_j^{bh}\left(v_j^{bh} + v_i^p + \left(\xi_j^{bh} + \xi_i^p\right)^2\right). \end{aligned} \quad \textbf{[S26]}$$

The approximation is exact when the point-spread function $\tilde{\mathcal{P}}_{bh}$ is precisely a single Gaussian, but is generally valid whenever $\tilde{\mathcal{P}}_{bh}$ is close to Gaussian (as is the case for the bead–handle system, where the corrections introduced by choosing $N_{bh} > 1$ are small). Eq. **S26** can be inverted to yield the desired parameter set $\{a_i^p, \xi_i^p, v_i^p\}$:

$$\begin{aligned} a_i^p &\approx a_i^{tot}, \\ \xi_i^p &\approx \xi_i^{tot} - \sum_{j=1}^{N_{bh}} a_j^{bh}\xi_j^{bh}, \\ v_i^p &\approx v_i^{tot} - \sum_{j=1}^{N_{bh}} a_j^{bh}\left[v_j^{bh} + \left(\xi_j^{bh} + \xi_i^{tot} - \sum_{k=1}^{N_{bh}} a_k^{bh}\xi_k^{bh}\right)^2\right], \end{aligned} \quad \textbf{[S27]}$$

where we have used the fact that $\sum_i^{N_{bh}} a_i^{bh} = 1$ due to normalization.

2. Experimental Verification of the Model for the Point-Spread Function

To check that the theoretical model of the point-spread function $\tilde{\mathcal{P}}_{bh}$ derived in *SI Text*, sections 1.4 and 1.5 is an accurate description of the handle and bead response in experiments, we analyzed control experiments on a system with only double-stranded DNA (dsDNA) handles and beads. Bead radii are R_b = 500 ± 25 nm, the trap strength is k_{trap} = 0.29 ± 0.02 pN/nm, and the handle parameters are extracted from the theoretical best fit described below. Four distinct experimental data sets are collected (Fig. S2): the first is from a pulling setup, where the trap separation is varied to give a trajectory of force F vs. total extension z (Fig. S2*A*, blue curve); the other data sets are trajectories of extension z as a function of time collected at three different constant trap separations. These three trajectories can be binned, and projected onto the constant-force ensemble using the same method (Eq. **S5**) as described above for the full system, yielding probability distributions $\tilde{\mathcal{P}}_{bh}(z;F_0)$ for the total end-to-end extension of the bead–handle system (Fig. S2 *B–D*, blue curves). The constant-force value F_0 for each projection is chosen equal to the mean force in each of the three trajectories, namely F_0 = 9.4 ± 0.7, 11.5 ± 0.8, and 12.7 ± 0.9 pN.

The experimental data were collected at 100 kHz, with no additional time averaging beyond the electronic filtering intrinsic to the detection and recording apparatus. Before the projection onto the constant force ensemble, the finite bandwidth scaling (FBS) method (*SI Text*, section 6) was used to approximately correct the raw experimental data for distortions due to electronic filtering and noise. In the absence of these corrections, the $\tilde{\mathcal{P}}_{bh}(z;F_0)$ from the raw data are given by the dashed curves in Fig. S2 *B–D*.

The SE margin (68% confidence interval) for each point in the $\tilde{\mathcal{P}}_{bh}$ distribution is marked by a light blue band, reflecting uncertainties in apparatus and FBS parameters, as well as statistical error due to sampling. Details of the error estimation procedure are in *SI Text*, section 7. The median SE in the z range shown varies from 3% to 5% between the three trajectories.

We use the theoretical model of *SI Text*, sections 1.4 and 1.5 to simultaneously fit all four experimental data sets with a single set of handle parameters, yielding best-fit values: L = 173 ± 2 nm, l_p 11 ± 1 nm, and γ = 520 ± 70 pN. The theory has excellent agreement with all of the experimental results, with median deviations in $\tilde{\mathcal{P}}_{bh}(z;F_0)$ for the z range shown in Fig. S2 *B–D* varying from 1% to 3% between the three trajectories, comparable to the SE margins. The comparison between theory and experiments firmly establishes the remarkable accuracy of our theory in quantitatively describing the bead–handle system.

3. Generalized Rouse Model

3.1. Hamiltonian and Exact Probability Distribution for the Generalized Rouse Model. The generalized Rouse model (GRM) (6), illustrated schematically in Fig. S1*B*, is a Gaussian chain with N monomers, connected by $N - 1$ harmonic springs with an average extension a. A conformation of the GRM is specified by the monomer positions $\mathbf{r}_i$, $i = 1,\ldots,N$. To get behavior reminiscent of hairpin unzipping, an additional harmonic bond potential $V(|\mathbf{r}_N - \mathbf{r}_1|)$ is added between the endpoints $\mathbf{r}_1$ and $\mathbf{r}_N$; the force due to this potential is nonzero only if the endpoint separation is within a cutoff distance c. Under a constant external tension $F_0\hat{\mathbf{z}}$ the GRM Hamiltonian is

$$\mathcal{H}_{GRM} = \frac{3k_BT}{2a^2}\sum_{i=1}^{N-1}(\mathbf{r}_{i+1} - \mathbf{r}_i)^2 + V(|\mathbf{r}_N - \mathbf{r}_1|) - F_0\hat{\mathbf{z}}\cdot(\mathbf{r}_N - \mathbf{r}_1), \quad \textbf{[S28]}$$

where $V(r) = kr^2\Theta(c - r) + kc^2\Theta(r - c)$, and Θ is the unit step function. We choose parameters $n = 18$, $F_0 = 2.9\ k_BT$/nm (11.9 pN), a = 1 nm, c = 12 nm, and $k = 0.09\ k_BT/\text{nm}^2$ (0.37 pN/nm).

If we write the end-to-end vector $\mathbf{r}_N - \mathbf{r}_1$ in cylindrical coordinates as (ρ, ϕ, z), the exact probability distribution for this vector in equilibrium under constant force $F_0\hat{\mathbf{z}}$ is given by

$$\tilde{\mathcal{P}}_{GRM}(\rho,\phi,z;F_0) = A_{GRM}\exp\left(-\frac{3(\rho^2+z^2)}{2a^2(N-1)} - \beta V\left(\sqrt{\rho^2+z^2}\right) + \beta F_0 z\right), \quad \textbf{[S29]}$$

where A_{GRM} is a normalization constant. This distribution, projected onto the (ρ, z) plane, is illustrated in Fig. 2*A*, *Upper*. The peak at small z corresponds to the "folded" hairpin state (F) with an intact endpoint bond, whereas the peak at larger z is the unfolded (U) state. Integrating $\tilde{\mathcal{P}}_{GRM}(\rho,\phi,z;F_0)$ over ρ and ϕ, one obtains the marginal probability $\tilde{\mathcal{P}}_{GRM}(z;F_0)$,

$$\tilde{\mathcal{P}}_{GRM}(z;F_0) = A'_{GRM}e^{-\frac{3z^2}{2a^2(N-1)}+\beta F_0 z} \cdot \begin{cases} e^{-\frac{c^2\left(3+2a^2\beta k(N-1)\right)-3z^2}{2a^2(N-1)}} - \dfrac{3e^{-\beta k z^2}}{3+2a^2\beta k(N-1)}\left(e^{-\left(c^2-z^2\right)\left(\beta k+\frac{3}{2a^2(N-1)}\right)}-1\right) & z \le c \\ e^{-\beta k c^2} & z > c, \end{cases} \quad \textbf{[S30]}$$

with normalization constant A'_{GRM}. $\tilde{\mathcal{P}}_{GRM}(z;F_0)$ is plotted in Fig. 2*A*, *Lower*.

3.2. Testing the GRM Deconvolution at Various Forces and Trap Strengths. In Fig. 3*B* we showed that the deconvolution results for the GRM are robust when varying the handle parameters. In

Fig. S3 we demonstrate that the same conclusion holds when either the force F_0 or the trap strength k_{trap} is varied.

4. Weighted Histogram Analysis Method: Combining Trajectories from Experimental Runs at Different Trap Separations

The weighted histogram analysis method (7) (WHAM) is a powerful tool in analyzing optical tweezer experiments. By combining trajectories generated at different trap separations z_{trap} (resulting in different force scales $\overline{F}$), one can sample the full extent of the protein free-energy landscape, and use WHAM to construct a single energy profile using all of the trajectory data, as previously done in ref. 8, and, in a related, but different manner, in ref. 9. In the context of our theory, WHAM modifies step i of our procedure, allowing us to derive the equilibrium probability $\tilde{\mathcal{P}}_{\mathrm{tot}}(z_{\mathrm{tot}};F_0)$ at constant force F_0 based on information from multiple experimental trajectories. Consider a set of M experimental runs, where the ith trajectory consists of n_i data points and has a trap separation $z_{\mathrm{trap}}^{(i)}$. Except for $z_{\mathrm{trap}}^{(i)}$, all other system parameters are kept the same between runs. For each run one can calculate the normalized histogram of total end-to-end distances z_{tot}, yielding a probability distribution $\mathcal{P}_{\mathrm{tot}}^{(i)}(z_{\mathrm{tot}})$. This distribution is related to $\mathcal{Q}(z_{\mathrm{tot}})$, the unbiased z_{tot} probability in the absence of a trapping potential or external force, through Eq. **S4**. Inverting that equation, we can write

$$\mathcal{Q}_{\mathrm{tot}}(z_{\mathrm{tot}}) \approx C_i^{-1} e^{\frac{1}{2}\beta k_z \left(z_{\mathrm{trap}}^{(i)} - z_{\mathrm{tot}}\right)^2} \mathcal{P}_{\mathrm{tot}}^{(i)}(z_{\mathrm{tot}}) \equiv e^{\beta(U_i(z_{\mathrm{tot}})-F_i)} \mathcal{P}_{\mathrm{tot}}^{(i)}(z_{\mathrm{tot}}), \quad \textbf{[S31]}$$

where C_i is a normalization constant, $U_i(z_{\mathrm{tot}}) = k_z(z_{\mathrm{trap}}^{(i)} - z_{\mathrm{tot}})^2/4$, and $F_i = \beta^{-1} \ln C_i$. In the case of one trajectory ($M = 1$), Eq. **S31** is a way to estimate $\mathcal{Q}_{\mathrm{tot}}(z_{\mathrm{tot}})$, from which one can calculate $\tilde{\mathcal{P}}_{\mathrm{tot}}(z_{\mathrm{tot}};F_0) = \exp(\beta F_0 z_{\mathrm{tot}})\mathcal{Q}_{\mathrm{tot}}(z_{\mathrm{tot}})$. This is just the standard step i procedure described earlier.

When $M > 1$, Eq. **S31** provides a different estimate of $\mathcal{Q}_{\mathrm{tot}}(z_{\mathrm{tot}})$ for each i, which ideally should be combined to give a single best approximation. The WHAM method resolves this problem, yielding a best estimate for $\mathcal{Q}_{\mathrm{tot}}(z_{\mathrm{tot}})$ of the form

$$\mathcal{Q}_{\mathrm{tot}}(z_{\mathrm{tot}}) = A \frac{\sum_{i=1}^{M} n_i \mathcal{P}_{\mathrm{tot}}^{(i)}(z_{\mathrm{tot}})}{\sum_{j=1}^{M} n_j e^{-\beta\left(U_j(z_{\mathrm{tot}})-F_j\right)}}, \quad \textbf{[S32]}$$

where A is a normalization constant. The unknown parameters $\{F_i\}$ are given by

$$F_i = -\frac{1}{\beta} \ln \left[\int dz_{\mathrm{tot}}\, \mathcal{Q}_{\mathrm{tot}}(z_{\mathrm{tot}}) e^{-\beta U_i(z_{\mathrm{tot}})} \right]. \quad \textbf{[S33]}$$

Eqs. **S32** and **S33** are a coupled system of equations for $\mathcal{Q}_{\mathrm{tot}}(z_{\mathrm{tot}})$ and $\{F_i\}$. We solve these by making an initial guess for the set $\{F_i\}$, substituting it into Eq. **S32** to find $\mathcal{Q}_{\mathrm{tot}}(z_{\mathrm{tot}})$, and using this estimate for $\mathcal{Q}_{\mathrm{tot}}(z_{\mathrm{tot}})$ in Eq. **S33** to find a new set of $\{F_i\}$. The process is iterated until we converge to a self-consistent solution to both equations. Once we have a best estimate of $\mathcal{Q}_{\mathrm{tot}}(z_{\mathrm{tot}})$, we can calculate $\tilde{\mathcal{P}}_{\mathrm{tot}}(z_{\mathrm{tot}};F_0)$ as above, completing step i of the construction.

5. Leucine Zipper Simulations

5.1. Self-Organized Polymer Model for the LZ26 Leucine Zipper. The amino acid sequence for a single α-helical strand of the LZ26 coiled coil is as follows (grouped into heptad repeats): MCQLEQK VEELLQK NYHLEQE VARLKQL VGELEQK VEELLQK NYHLEQE VARLKQL VGELEQK VEELLQK NYHLEQE VARLKQL VGEC. The sequence is the same as in ref. 10, except that we have left out 4 residues at the beginning and 3 from the final heptad, for a total of 88 residues per strand. As in the experiment (10), the handles are attached at the cysteine in position b of the first heptad, and the cross-linking between strands is at the cysteine in position d of the last heptad. (For consistency, when comparing simulations with or without the handles/traps, end-to-end distance for the protein is always measured between the two N-terminal cysteines.) Although the crystal structure is not available for LZ26, it is believed to be similar to three GCN4 leucine zipper domains (Protein Data Bank ID code 2ZTA) (11) in series. Thus, we constructed a model for the native structure based on GCN4, connecting the leucine zipper segments in such a way that the distances between neighboring C_α positions and angles of superhelical coiling formed a continuous pattern as one moves along LZ26.

Going from N- to C terminus on one strand and returning C- to N terminus on the other, let us label the residues $i = 1,\ldots,N_{\mathrm{res}}$, where $N_{\mathrm{res}} = 176$, where $1 \le i \le 88$ corresponds to one strand, and $88 < i \le 176$ corresponds to the other. Every nonneighboring pair of residues (i, j), where $|i - j| > 1$, is assigned to one of three sets: $\mathcal{S}$ (secondary structure pairs), $\mathcal{T}$ (tertiary structure pairs), and $\mathcal{R}$ (remainder). The set $\mathcal{S}$ consists of all pairs where $|i - j| = 4$ and i, j share the same strand, representing residues interacting through α-helical hydrogen bonding. The set $\mathcal{T}$ consists of all pairs where i, j are on different strands, and the distance between the two residues in the native structure $r_{i,j}^0$ is below a cutoff: $r_{i,j}^0 < R_c = 0.8$ nm. These pairs are involved in tertiary interactions between the two α-helical coils. All other nonneighboring pairs that do not satisfy the criteria for $\mathcal{S}$ or $\mathcal{T}$ fall into the set $\mathcal{R}$, and only interact via repulsive Lennard-Jones potentials.

The variant self-organized polymer (SOP) Hamiltonian for LZ26 has the form

$$\begin{aligned}
\mathcal{H}_{\mathrm{SOP}} = {} & \frac{k_{\mathrm{bond}}}{2} \sum_{i=1}^{N_{\mathrm{res}}-1} \left(r_{i,i+1} - r_{i,i+1}^0\right)^2 \\
& + \frac{k_{\mathrm{ang}}}{2} \sum_{i=1}^{N_{\mathrm{res}}-2} \left[1 - \cos\left(\theta_{i,i+1,i+2} - \theta_0\right)\right] \Delta_{i,i+1,i+2} \\
& - \sum_{(i,j)\in\mathcal{S}} \epsilon_{\mathrm{hb}} \Big[1 + k_{\mathrm{bond}}^{\mathrm{hb}} \left(r_{i,j} - r_{i,j}^0\right)^2 + k_{\mathrm{ang}}^{\mathrm{hb}} \Big(\left(\theta_{i,j,j-1} - \theta_{i,j,j-1}^0\right)^2 \\
& + \left(\theta_{i+1,i,j} - \theta_{i+1,i,j}^0\right)^2 + \left(\phi_{i+1,i,j,j-1} - \phi_{i+1,i,j,j-1}^0\right)^2 \Big) \Big]^{-1} \\
& + \sum_{(i,j)\in\mathcal{T}} \chi \left| \epsilon_{\mathrm{BT}}(i,j) - \epsilon_s \right| V_{\mathrm{LJ}} \left(\frac{r_{i,j}^0}{r_{i,j}} \right) + \sum_{(i,j)\in\mathcal{R}} \epsilon_{\mathrm{rep}} \left(\frac{\sigma}{r_{i,j}} \right)^6 .
\end{aligned} \quad \textbf{[S34]}$$

The first term is the nearest-neighbor bond potential, where $r_{i,j}$ is the distance between residues i and j, and the spring constant $k_{\mathrm{bond}} = 200$ kcal/mol · nm^2. The second term is the bond angle potential, with the spring constant $k_{\mathrm{ang}} = 2$ kcal/mol. The angle between the bonds (j, i) and (j, k) is $\theta_{i,j,k}$, and the equilibrium value $\theta_0 = 0.583\pi$ rad $= 105°$, a typical bond angle in protein structures (12). The factor $\Delta_{i,j,k} = 1$ if i, j, k are all on the same strand, 0 otherwise. The relative softness of the bond angle potential, together with the form of the secondary structure interactions detailed below, ensures that the two strands in the unfolded LZ26 (with all interstrand tertiary contacts broken) have a persistence length of ~0.7 nm, consistent with experimental measurements (10).

The third term in Eq. **S34** accounts for the effects of hydrogen bonding along the α-helical backbone, and is based on a similar form developed for RNA (13). We mimic the directionality de-

pendence of hydrogen bonds by making the bond energy depend not only on the distance $r_{i,j}$, but also on bond and dihedral angles defined by the four residues $i+1$, i, j, and $j-1$, with $|i-j|=4$. For each (i,j) there are two bond angles $\theta_{i,j,j-1}$ and $\theta_{i+1,i,j}$ and one dihedral angle $\phi_{i+1,i,j,j-1}$. The equilibrium values of the angles, denoted by a superscript 0, are calculated from the corresponding quantities in the native structure. Only when the distance, bond angles, and dihedral angles are all simultaneously equal to the equilibrium values does the hydrogen bond potential reach its energy minimum $-\epsilon_{\rm hb}$, where $\epsilon_{\rm hb}>0$. Thus, the minimum is reached only when the entire $(i,i+4)$ strand segment adopts a structure resembling a single α-helical turn. The α-helical propensity of an $(i,i+4)$ segment is determined by the energy scale $\epsilon_{\rm hb}$ and the sensitivity parameters $k^{\rm hb}_{\rm bond}$, $k^{\rm hb}_{\rm ang}$. Larger values for the sensitivity parameters increase the brittleness of the α-helix, making it more likely to be destabilized due to thermal fluctuations. To calibrate the parameters, we define a helix function $H(i,j)$ for any $(i,j)\in\mathcal{S}$,

$$H(i,j)=\sqrt{\left(\frac{r_{i,j}}{r^0_{i,j}}-1\right)^2+\left(\frac{\theta_{i,j,j-1}}{\theta^0_{i,j,j-1}}-1\right)^2+\left(\frac{\theta_{i+1,i,j}}{\theta^0_{i+1,i,j}}-1\right)^2+\left(\frac{\phi_{i+1,i,j,j-1}}{\phi^0_{i+1,i,j,j-1}}-1\right)^2},$$

[S35]

reflecting the rms deviation of the bond distances and angles from their equilibrium values. We use $H(i,j)$ as a measure of helix content, by counting the fraction of pairs in $\mathcal{S}$ where $H(i,j)$ is less than a cutoff $H_c=0.5$. It is known from thermal denaturation experiments on GCN4 (14) that the individual α-helices upon unzipping are unstable, with ~17% helical content. In contrast, the tertiary contacts in the coiled-coil structure stabilize helix formation, resulting in a much higher helical content of ~81%. We expect qualitatively similar behavior in LZ26 in the case of force denaturation, and thus tune the sensitivity parameters to yield a large difference in the helix content between the unfolded and folded states. The parameter values are set at $\epsilon_{\rm hb}=3.85$ kcal/mol, $k^{\rm hb}_{\rm bond}=10\ {\rm nm}^{-2}$, and $k^{\rm hb}_{\rm ang}=40\ {\rm rad}^{-2}$. For these values we find a helix content of 3 and 82%, respectively, for the unfolded and folded states of LZ26 under a constant force of $F_0=12.3$ pN.

The fourth term in Eq. **S34** describes tertiary interactions between the two strands of LZ26, $(i,j)\in\mathcal{T}$. These have a residue-dependent energy $\chi|\epsilon_{\rm BT}(i,j)-\epsilon_s|$. Here, χ is an overall prefactor, $\epsilon_{\rm BT}(i,j)$ is the Betancourt–Thirumalai (BT) contact energy for residues i and j (15), and ϵ_s shifts the zero of the energy scale (16). To get a leucine zipper that unfolds at the experimental force scale of ~12 pN, we choose $\chi=2.25$ and energy shift $\epsilon_s=0.7\,k_BT$. The tertiary interactions use a modified Lennard-Jones potential of the form

$$V_{\rm LJ}(x)=\begin{cases}x^6-\frac{3}{2}x^4-\frac{1}{2} & x\le 1\\ x^{12}-2x^6 & x>1\end{cases}.$$

[S36]

This has the standard 12–6 form at large distances, but a softer short-range repulsive core, increasing with the inverse 6th rather than 12th power. The choice of the softer potential is made to allow for a longer simulation time step, although not having a significant impact on the large-scale dynamics of the system (17).

The final term in Eq. **S34** describes purely repulsive interactions among the remaining nonneighboring pairs $(i,j)\in\mathcal{R}$, with energy factor $\epsilon_{\rm rep}=1$ kcal/mol and range $\sigma=0.38$ nm. We use the inverse sixth power in the repulsive potential for the same reasons as above.

5.2. Semiflexible Bead–Spring Model for the DNA Handles. Each dsDNA handle is modeled as a chain of $N_{\rm h}$ beads of radius $a=1$ nm, corresponding to a contour length $L=2aN_{\rm h}$. The handle Hamiltonian is

$$\mathcal{H}_{\rm h}=\frac{k_{\rm bond}}{2}\sum_{i=1}^{N_{\rm h}-1}\left(r_{i,i+1}-2a\right)^2+\frac{l_pk_BT}{2a}\sum_{i=1}^{N_{\rm h}-2}\left[1-\cos(\theta_{i,i+1,i+2})\right],$$

[S37]

where $r_{i,i+1}$ are the distances between neighboring beads, $k_{\rm bond}=200$ kcal/mol · nm^2, l_p is the persistence length, and $\theta_{i,i+1,i+2}$ are angles between consecutive bonds. The two terms are stretching and bending energies, respectively. The handle elastic modulus $\gamma=2ak_{\rm bond}=2{,}780$ pN. For the persistence length we consider, $l_p=20$ nm, at the applied tension due to the traps, the handles (and unfolded portions of the protein) are almost fully extended, and there is negligible probability of the chain overlapping itself or protein residues in the vicinity of the handle attachment point. Hence, there is no need to include excluded volume interactions for the handles. The covalent linkers that attach the handles either to the cysteine residue at the protein N terminus or a point on the bead surface are modeled as simple harmonic springs with strength $\kappa=k_{\rm bond}$ and length $\ell=1.5$ nm.

5.3. Simulation Time Scales. Let $\mu_0=1/6\pi\eta a$ be the mobility of a sphere of radius $a=1$ nm, where $\eta=0.89$ mPa · s is the viscosity of water at $T=298$ K. This will be the mobility of our DNA handle beads, whereas for the large polystyrene beads the corresponding mobility is $\mu_b=\mu_0a/R_b$. The rotational diffusion of the polystyrene bead is characterized by a mobility $\mu^r_{\rm b}=3\mu_0a/4R_b^3$. For the protein residues we choose a mobility $\mu_{\rm res}=3.36\mu_0$, corresponding to an effective hydrodynamic radius of 0.30 nm. The characteristic Brownian dynamics time scale associated with μ_0 is $\tau_0=a^2/\mu_0k_BT=4$ ns. To avoid numerical errors, our simulation time step τ should be a small fraction of τ_0, and we obtained reliable results using $\tau=5\times10^{-6}\ \tau_0=0.02$ ps. For LZ26 both with and without handles/beads, we ran ~260 long trajectories at various force conditions (or trap separations), totaling to $\sim10^{12}$ simulation time steps, or 20 ms, with data collection every 10^4 steps. (In the case of the simulations involving the GRM hairpin instead of the protein, the time step $\tau=3\times10^{-4}\ \tau_0=1.2$ ps, and the total trajectory data for each GRM parameter set corresponded to 160–180 ms.)

6. FBS: Correcting for the Effects of Electronic Filtering, Time Averaging, and Noise

Before the data from optical tweezer experiments can be used to reconstruct the intrinsic biomolecule free-energy landscape, one must consider the inevitable distortions due to noise, the electronic systems involved in data recording, and any additional filtering done as part of the collection protocol. We have developed a method (FBS) to correct for these distortions. In the following we first derive the basic FBS scaling relations, and then verify them using both simulation and experimental data sets.

6.1. FBS Theory. Understanding how the time series of bead positions is distorted as part of the measurement process requires a detailed spectral analysis of all components in the dual optical tweezer apparatus. The spectral properties of the experimental system used to collect the data in our work have been extensively characterized by von Hansen et al. (18), allowing us to develop a simplified theory which approximates the most important sources of distortion. Our theory fits all of the experimental data sets under consideration, but it can be easily modified to include additional complications that we ignore (for example, cross-talk between the two laser traps) as well as the details of other experimental setups.

Let $z^{\rm raw}_{\rm tot}(t)$ be the trajectory of bead–bead separations along the $\hat{\mathbf{z}}$ axis recorded during the experiment. This raw data set is based on the signal from the silicon photodiode devices that

measure the deflection of the lasers due to bead displacements. This output is then processed and amplified by the electronic system used in the recording apparatus. If $z_{\rm tot}(t)$ is the actual trajectory of bead displacements, inaccessible to the experimentalist, the recorded output $z_{\rm tot}^{\rm raw}(t)$ is related to $z_{\rm tot}(t)$ as

$$z_{\rm tot}^{\rm raw}(t) = \int_{-\infty}^{t} f(t-t')\left(z_{\rm tot}(t') + \eta(t')\right). \qquad \textbf{[S38]}$$

The deviation of $z_{\rm tot}^{\rm raw}(t)$ from $z_{\rm tot}(t)$ stems from two main effects: (*i*) an additive noise component $\eta(t)$, which includes environmental noise such as vibrations of the optical elements in the apparatus and electronic noise in the detectors (18). For simplicity, we model the noise as Gaussian white noise with zero mean and variance equal to ν: $\langle\eta(t)\rangle = 0$, $\langle\eta(t')\rangle = \nu\delta(t-t')$, and $\langle\eta(t)z_{\rm tot}(t')\rangle = 0$, where $\langle\rangle$ denotes an equilibrium ensemble average; (*ii*) convolution with a kernel function $f(t)$, which reflects the filtering properties of the photodiodes and electronics. Any additional time averaging or filtering carried out by the experimentalist on the recorded data series will be considered explicitly later on, and is not included in $f(t)$. The analysis of ref. 18. yielded the following form for the filter kernel in the frequency domain:

$$f(\omega) = \left[\lambda + \frac{1-\lambda}{1-i\omega\tau_{pf}}\right]\frac{1}{B_8\left(i\omega\tau_{bf}\right)}, \qquad \textbf{[S39]}$$

where for our laser optical tweezer setup $\lambda = 0.6 \pm 0.05$, $\tau_{pf} = 6 + 1$ μs, $\tau_{bf} = 5$ μs, and $B_8(x)$ is the eighth-order Butterworth polynomial. The term in the square brackets above originates in a physical phenomenon known as "parasitic filtering" (19, 20), arising from the transparency of the silicon in the photodiode to the laser light with wavelength 1,064 nm used in the experiment: a fraction $1 - \lambda$ of the photocurrent from the detector is produced with a lag time τ_{pf} relative to the photon signal. The second term in Eq. **S39**, involving the Butterworth polynomial, is due to the subsequent electronic amplification of the signal from the detector, which acts like a Butterworth low-pass filter with characteristic time scale τ_{bf}, such that at the frequency $\omega = \tau_{bf}^{-1}$ the signal amplitude is attenuated by 3 dB. Because the form of Eq. **S39** is too complicated for use in our analytical theory, we will approximate $f(t)$ as a generic first-order low-pass filter, exploiting the fact that both the parasitic and electronic terms act to attenuate high-frequency portions of the signal,

$$f(\omega) \approx \frac{1}{1-i\omega\tau_f}, \qquad \textbf{[S40]}$$

where $\tau_f = 7$ μs. This effective filtering time scale τ_f is derived by demanding that Eq. **S40** exhibit the same degree of attenuation at $\omega = \tau_{bf}^{-1}$ as Eq. **S39**.

Although these distortions are expressed in the frequency domain, they have observable consequences for the equilibrium probability distribution of bead–bead separations. As an example, consider the raw autocorrelation function $C_{\rm raw}(t) = \langle(z_{\rm tot}^{\rm raw}(t) - \bar{z}_{\rm tot}^{\rm raw})(z_{\rm tot}^{\rm raw}(0) - \bar{z}_{\rm tot}^{\rm raw})\rangle$, where $\bar{z}_{\rm tot}^{\rm raw}$ is the mean recorded bead–bead separation. The variance of the raw probability distribution $\mathcal{P}_{\rm tot}^{\rm raw}(z_{\rm tot}^{\rm raw})$ is equal to $C_{\rm raw}(0)$. From Eqs. **S38** and **S40**, the raw autocorrelation is related to the true one $C(t) = \langle(z_{\rm tot}(t) - \bar{z}_{\rm tot})(z_{\rm tot}(0) - \bar{z}_{\rm tot})\rangle$ by

$$C_{\rm raw}(t) = \frac{\nu}{2\tau_f}e^{-t/\tau_f} + \int_{-\infty}^{\infty} dt' \frac{e^{-|t-t'|/\tau_f}}{2\tau_f} C(t'). \qquad \textbf{[S41]}$$

The first term in Eq. **S41**, due to noise, tends to increase the variance $C_{\rm raw}(0)$ relative to $C(0)$. The second term, due to filtering, is always less than $C(0)$ because it is an average over $C(t')$, and $C(t' \neq 0) < C(0)$. Noise broadens the measured distribution and filtering narrows it. However, without knowing the amplitude of the noise ν, it is unclear whether the filtering due to the detectors and electronics under- or overcompensates for the noise, and how far $\mathcal{P}_{\rm tot}^{\rm raw}(z_{\rm tot}^{\rm raw})$ deviates from the true distribution $\mathcal{P}_{\rm tot}(z_{\rm tot})$. Thus, we need a way to estimate ν.

The situation is even more complicated because the experimentalist may choose to apply additional filtering on the recorded data, for example as a way of manually removing noise and unwanted high-frequency components of the signal (because the dynamics of interest typically occurs at frequencies much lower than imposed filter cutoff). For the GCN4 leucine zipper, the data sets recorded at 100 kHz (corresponding to a sampling time step $\tau_s = 10$ μs) were subsequently filtered in real time during collection by averaging every five consecutive time steps together. Such averaging acts like a low-pass filter, and so has narrowing effects on the equilibrium probability distribution qualitatively similar to the filtering described above. Some type of additional filtering of this kind is a common experimental practice (see refs. 21–23 for recent examples, involving either an averaging or eight-pole Bessel filter). It turns out, however, that we can take advantage of the filtering protocol: by varying the degree of filtering we will use it to approximately extrapolate features of the true probability distribution.

Let us concentrate on the simple case of filtering the recorded data by averaging every n consecutive points into a single value. If the collection time step is τ_s, the original raw data are represented by the recorded time series $\{z_{\rm tot}^{\rm raw}(t_j)\}$, where $t_j \equiv j\tau_s$ for $j = 1,2,3,\ldots$ The averaged data are a time series $\{z_{\rm tot}^{{\rm raw},n}(t_{nj})\}$, where $z_{\rm tot}^{{\rm raw},n}(t_{nj}) = n^{-1}\sum_{i=nj-n+1}^{i=nj} z_{\rm tot}^{\rm raw}(t_i)$. For the averaged time series we will focus on two quantities, both related directly to its autocorrelation $C_{{\rm raw},n}(t)$: the variance $\langle(z_{\rm tot}^{{\rm raw},n} - \bar{z}_{\rm tot}^{{\rm raw},n})^2\rangle = C_{{\rm raw},n}(0)$, and the mean-squared displacement (MSD) between consecutive points $\langle(z_{\rm tot}^{{\rm raw},n}(n\tau_s) - z_{\rm tot}^{{\rm raw},n}(0))^2\rangle = 2(C_{{\rm raw},n}(0) - C_{{\rm raw},n}(n\tau_s)) \equiv \Delta_{{\rm raw},n}(n\tau_s)$. In a more complicated fashion, these two quantities can also be expressed in terms of the original autocorrelation $C_{\rm raw,1}(t) = C_{\rm raw}(t)$ before averaging:

$$C_{{\rm raw},n}(0) = \frac{1}{n}C_{\rm raw}(0) + \frac{2}{n^2}\sum_{j=1}^{n-1}(n-j)C_{\rm raw}(j\tau_s),$$

$$\Delta_{{\rm raw},n}(n\tau_s) = \frac{2}{n}C_{\rm raw}(0) + \frac{2}{n^2}\sum_{j=1}^{n}(2n-3j)C_{\rm raw}(j\tau_s) - \frac{2}{n^2}\sum_{j=1}^{n-1}(n-j)C_{\rm raw}((n+j)\tau_s). \qquad \textbf{[S42]}$$

We know that $C_{\rm raw}(t)$ is related to the unknown true correlation $C(t)$ through Eq. **S41**, so we can complete the theoretical description by specifying a form for $C(t)$. A generic correlation function can be expanded as a sum of exponentials $C(t) = \sum_{i=1}^{\infty} A_i \exp(-t/\tau_i)$, with relaxation times $\tau_1 < \tau_2 < \ldots$ We will be interested in correlations on the shortest accessible time scales, $t \sim \mathcal{O}(\tau_s)$, so we plug the expression for $C(t)$ into Eq. **S41** and expand for small t, keeping the contribution from the τ_1 exponential and lowest-order corrections from the $\tau_{i>1}$ terms:

$$C_{\rm raw}(t) = \frac{\nu}{2\tau_f}e^{-t/\tau_f} + \sum_{i=1}^{\infty}\frac{A_i\tau_i}{\tau_i^2 - \tau_f^2}\left(\tau_i e^{-t/\tau_i} - \tau_f e^{-t/\tau_f}\right)$$

$$\approx \frac{\nu}{2\tau_f}e^{-t/\tau_f} + \frac{A_1\tau_i}{\tau_1^2 - \tau_f^2}\left(\tau_1 e^{-t/\tau_1} - \tau_f e^{-t/\tau_f}\right) + A_c - B_c\left(t + t_f e^{-t/\tau_f}\right), \qquad \textbf{[S43]}$$

where $A_c = \sum_{i=2}^{\infty} A_i$, $B_c = A_2/\tau_2$. If necessary, the expansion can be extended to higher orders, but the above form was sufficient

to fit all of the simulation and experimental cases which we analyze below.

Eqs. **S42** and **S43** completely define the variance $C_{\mathrm{raw},n}(0)$ and MSD $\Delta_{\mathrm{raw},n}(n\tau_s)$ in terms of five unknown parameters: ν, A_1, τ_1, A_c, and B_c. By averaging the recorded time series $\{z_{\mathrm{tot}}^{\mathrm{raw}}(t_j)\}$ for different values of n (varying the effective filter bandwidth), we construct curves of $C_{\mathrm{raw},n}(0)$ and MSD $\Delta_{\mathrm{raw},n}(n\tau_s)$ as a function of n. Fitting these curves to Eqs. **S42** and **S43**, we can then extract the unknown parameters. This allows us to estimate the true variance of the probability distribution,

$$C(0) = A_1 + A_c. \qquad \textbf{[S44]}$$

Because we are using properties of time series at different effective bandwidths to gain information about the true, "infinite" bandwidth limit, we call our method *finite bandwidth scaling* (FBS). The analogy is to finite size scaling (24), where thermodynamic properties of systems on finite lattices are extrapolated to the infinite lattice limit. One of the nice features of FBS is that the scaling analysis can be carried out even when we can only calculate $C_{\mathrm{raw},n}(0)$ and $\Delta_{\mathrm{raw},n}(n\tau_s)$ for a subset of n values. For example, in the leucine zipper case below, the available time series corresponds to $n = 5$ because the data were time averaged during collection. From the $n = 5$ data we can construct trajectories for $n = 10, 15, 20, \ldots$ This subset is sufficient for the FBS extrapolation.

Once we know $C(0)$, how can we use it to approximately reconstruct the true distribution $\mathcal{P}_{\mathrm{tot}}(z_{\mathrm{tot}})$? Keep in mind that the variance $C(0) = \langle z_{\mathrm{tot}}^2 \rangle - \langle z_{\mathrm{tot}} \rangle^2 = \int dz (z - \langle z_{\mathrm{tot}} \rangle)^2 \mathcal{P}_{\mathrm{tot}}(z)$. The simplest estimate for $\mathcal{P}_{\mathrm{tot}}(z_{\mathrm{tot}})$ is to start with the measured, averaged distribution $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},n}$ for some n, and deform it in one of two ways, changing its variance by an amount $\delta_C = |C(0) - C_{\mathrm{raw},n}(0)|$:

i) If $C_{\mathrm{raw},n}(0) < C(0)$, we carry out a convolution with a normalized Gaussian of variance δ_C,

$$\mathcal{P}_{\mathrm{tot}}(z_{\mathrm{tot}}) \approx \int_{-\infty}^{\infty} dz \mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},n}(z_{\mathrm{tot}} - z) \frac{e^{-z^2/2\delta_c}}{\sqrt{2\pi\delta_C}}. \qquad \textbf{[S45]}$$

ii) If $C_{\mathrm{raw},n}(0) > C(0)$, we do a deconvolution instead, solving

$$\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},n}(z_{\mathrm{tot}}) \approx \int_{-\infty}^{\infty} dz \mathcal{P}_{\mathrm{tot}}(z_{\mathrm{tot}} - z) \frac{e^{-z^2/2\delta_c}}{\sqrt{2\pi\delta_C}} \qquad \textbf{[S46]}$$

for $\mathcal{P}_{\mathrm{tot}}$. The latter can be carried out using the numerical deconvolution technique described in *SI Text*, section 1.6. After the deformation, the estimated $\mathcal{P}_{\mathrm{tot}}$ will, by construction, have the correct variance $C(0)$. We should recover roughly the same $\mathcal{P}_{\mathrm{tot}}$ starting from $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},n}$ for any n in the range where the FBS scaling is valid, as we will demonstrate in the examples below. In systems with multiple states, where there is more than one peak in the measured distribution, it is more accurate to carry out the FBS analysis separately on each state, and apply the corresponding specific deformation for each peak. This can be done with the aid of hidden Markov model (HMM, ref. 25) partitioning of the time series, as described in the next section for the case of the GRM and leucine zipper.

The FBS method has several limitations: (*i*) using a Gaussian to deform $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},n}$ into $\mathcal{P}_{\mathrm{tot}}$ is an assumption, because all we strictly know that the actual point-spread function is the variance δ_C. The smaller the variance, the more valid the assumption, because the potential non-Gaussian contributions to the point-spread function become less significant. We can also test the assumption from our measured data by checking whether the $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},n}$ for various n can be mapped to each other by Gaussian deformations. For all of the systems analyzed in our work this is indeed the case. (*ii*) Gaussian deformations map individual peaks into slightly broader or narrower peaks, but do not produce new peaks. Hence, if there is a state with a very short lifetime that is smeared out by the filtering (either the parasitic, electronic, or additional filtering), yielding no distinct peak in $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},n}$, the FBS method will not be able to reconstruct its properties. Whether or not the experimentalist chooses to do additional averaging, the intrinsic time resolution τ_f of the apparatus puts fundamental constraints on what we can learn from the measured time series. Transitions occurring on time scales faster than τ_f will be lost to us. (*iii*) In a similar way, the characteristic relaxation times of the trapped beads also impose limits. To illustrate this, take two protein states S_1 and S_2, which have a small difference in their mean end-to-end distance along the force direction, and assume S_1 is only accessible from S_2. If the mean lifetime of S_1 is much smaller than S_2, such that it is shorter than the bead relaxation time, transitions such as $S_2 \rightarrow S_1 \rightarrow S_2$ will correspond only to negligible excursions in the measured trajectory of bead displacements, because the beads do not have enough time to relax to the equilibrium position associated with S_1 before the protein returns to S_2. If the mean-squared distance of the excursions is smaller than the noise amplitude in the recorded time series, the existence of state S_1 will be hidden from the experimentalist, regardless of the apparatus filtering time scale τ_f. In summary, the distribution produced by FBS is an approximation to the truth: the method can correct distortions produced by noise and filtering, but it only works for states in the energy landscape which leave some signature of themselves in the measured time series.

6.2. Testing FBS on Simulation and Experimental Data. As a first test of the FBS theory, we analyze a Brownian dynamics simulation trajectory of the GRM model in an optical tweezer setup (Fig. 4). The trap separation $z_{\mathrm{trap}} = 1{,}298$ nm; all of the other parameters are listed in Table S1. A computer simulation has perfect recording of data, with no environmental noise or apparatus filtering effects, hence it can test the FBS theory of Eqs. **S42** and **S43** in the limit $\nu = \tau_f = 0$. In this case the true distribution is just the $n = 1$ raw distribution $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw}} = \mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},1}$, plotted in Fig. S4*A* (gray curve). If the FBS scaling is valid, we should be able to map any distribution for $n > 1$ onto the $n = 1$ result by applying the FBS correction procedure described above.

The GRM model exhibits two states, native N and unfolded U, which have distinct dynamical properties. Hence, it is more accurate to apply the FBS method separately to just those portions of the time series belonging to each state. Partitioning the time series by state requires estimating the most likely sequence of states that corresponds to the data. HMM (25) is a general tool for this task. The probability distribution can be accurately decomposed into Gaussians corresponding to each state, as depicted in Fig. S4*A*, which define likelihoods for any observed z_{tot} data point in the trajectory to belong to one or the other state. We then process the entire trajectory through the Baum–Welch algorithm (26) to find optimal values for the unknown transition probabilities between states, and finally construct the most likely state sequence using the Viterbi algorithm (27). Fig. S4*B* shows a fragment of the trajectory, colored according to the state assignment resulting from HMM.

The variance $C_{\mathrm{raw},n}(0)$ and MSD $\Delta_{\mathrm{raw},n}(n\tau_s)$ are then calculated as a function of n from the trajectory fragments belonging to a certain state. For a given n, the calculation involves averaging over data points within a time window up to $2n\tau_s$ in length, so getting good statistics requires having many fragments longer

than $2n\tau_s$. This will be true so long as $2n\tau_s$ is much smaller than the mean lifetime of the state, putting a practical upper bound on n. Fig. S4 *C* and *D* plots the results for $C_{\mathrm{raw},n}(0)$ and $\Delta_{\mathrm{raw},n}(n\tau_s)$, respectively (blue points: N state; red points: U state). Bars represent statistical SE due to finite sample size, as determined through jackknife estimation (28). The solid curves are fits to Eqs. **S42** and **S43**, from which we extract the FBS parameter values listed in Table S2.

With these values in hand, we can carry out the correction procedure: Fig. 4*D* shows the raw distributions $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},n}$ for $n = 1$, 50, 100, 150 (solid curves), and *E* shows the same distributions after they have been corrected according to the method outlined above ($n = 1$ needs no correction, but is included for comparison). The greater the degree of averaging (increasing n), the narrower the peaks in $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},n}$. However, the FBS method compensates for this, and all of the distributions in *E* have collapsed onto a single estimate for the true $\mathcal{P}_{\mathrm{tot}}$. As expected, this estimate agrees very well with the $n = 1$ result $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw}}$ (cyan points).

The second test of the FBS theory is on experimental data for a system with only dsDNA handles and beads, discussed in *SI Text*, section 2. FBS results for three different trajectories (corresponding to three values of the mean force F_0) are presented in Fig. S5. These data sets were recorded with a sampling rate of 100 kHz ($\tau_s = 10$ μs), with no additional averaging beyond the unavoidable filtering effects of the detectors and electronics. As a consequence of these effects, $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},1}$ is not the same as the true distribution, and the deviation grows larger as n is increased. The FBS best-fit results are shown in Table S2. In the fitting the noise amplitude ν is constrained to be the same among all three trajectories, because they are all collected on the same equipment. As in the previous example, $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},n}$ for various n can all be collapsed onto a single estimate for $\mathcal{P}_{\mathrm{tot}}$ through the FBS method. In Fig. S5*E* this estimate (solid curves) is compared with $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},1}$ (dashed curves) to emphasize that the distortions due to apparatus filtering are small but noticeable.

The final test is on the GCN4 leucine zipper experimental time series (trajectory 1, with parameters described in Table S1). As mentioned earlier, here we only can construct averaged data sets for $n = 5$, 10, 15, ... The $n = 1$ trajectory, at the original $\tau_s = 10$-μs sampling interval, is not available. Despite this limitation, the FBS scaling analysis works nicely, with results summarized in Fig. S6 and Table S2. We took advantage of the fact that the handle-only data sets, collected with the same optical tweezer setup as the leucine zipper (except with no protein), had direct information about $n = 1$ time scales, and thus probed higher frequencies than were accessible in the leucine zipper data. Because going to higher frequencies gives us better estimates of the background noise, we set the noise amplitude ν in the leucine zipper case to the best-fit value from the handle-only analysis. All other FBS parameters were fit individually for each state (I1, I2, and U) in the leucine zipper distribution. With FBS corrections, $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},n}$ for n going up to 25 (effective bandwidths as low as 4 kHz) all collapse onto a single estimate of the true $\mathcal{P}_{\mathrm{tot}}$.

7. Estimating Uncertainties in the Free-Energy Reconstruction

The free-energy reconstruction is only as good as the data on which it is based: the recorded time series which is the input, and the information about the apparatus which is used to analyze the time series and predict the intrinsic landscape. Both of these are subject to uncertainties, which will propagate into the final result. Let us first consider statistical uncertainties due to the finite length of the trajectories from which the input probability distribution $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw}}$ is determined. One of the advantages of the double optical trap setup is that it is exceptionally stable, allowing for data collection over periods >100 s. In the case of the leucine zipper, the slowest transition (from U to I2) occurs on time scales of 0.4–0.6 s in the force range of interest, so even a single trajectory contains $\sim\mathcal{O}(10^2)$ of the rarest observed conformational changes.

Thus, the distribution $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw}}$ has small statistical uncertainties. To quantitatively estimate the error, we use a block bootstrap method (29, 30) in the following manner: The trajectory is divided into blocks of length larger than the longest autocorrelation time, and a synthetic data set of the same length is generated by sampling with replacement from this set of blocks. Using a large number of synthetic data sets (>500) we can determine confidence intervals for each point in the $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw}}$ distribution. The number of blocks is varied until convergence is achieved in the error estimate. The results are shown in Fig. S7 *A* and *B* for two leucine zipper experimental trajectories (parameters as in Table S1). The $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw}}$ distributions (black curves) are surrounded by dark blue bands which represent the 68% confidence interval, or SE margin. The median error in the z range where $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw}} > 10^{-6}$, is 10% and 19%, respectively, for the two trajectories.

In reconstructing the intrinsic free-energy landscape $\tilde{\mathcal{F}}_{\mathrm{p}}$, this statistical error is compounded by uncertainties in all of the apparatus parameters that are used in the analysis: bead radii, trap strengths, handle properties, as listed in Table S1, as well as uncertainties in the FBS parameters used to correct the raw distributions (Table S2). We perform a Monte Carlo error estimate by sampling from Gaussian distributions of these parameters with SDs given by the uncertainties, and for each parameter set performing the complete free-energy reconstruction on the entire ensemble of synthetic data sets generated by the block bootstrap. To analyze the shape differences among the reconstructed landscapes, every $\tilde{\mathcal{F}}_{\mathrm{p}}$ is projected to the midpoint value of F_0, where the probabilities of states I1 and U are equal. ($F_0 = 12.3 \pm 0.9$ pN and 12.1 ± 0.9 pN from trajectories 1 and 2, respectively.) Although computationally intensive, this procedure allows us to estimate 68% confidence intervals for $\tilde{\mathcal{F}}_{\mathrm{p}}$, shown as light blue bands for the two trajectories in Fig. S7 *C* and *D*. For comparison, if one assumed no uncertainty in the apparatus parameters, one would get the much narrower dark blue bands, representing just the error in $\tilde{\mathcal{F}}_{\mathrm{p}}$ from the finite sampling of $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw}}$. Clearly, the uncertainties in the apparatus parameters are the predominant source of error.

With both apparatus and sampling uncertainties included, the median SE over the z range where $\tilde{\mathcal{F}}_{\mathrm{p}} < -k_BT\ln\ (10^{-6}) \approx 14k_BT$ is 10% in both trajectories. This corresponds to $\sim 0.4k_BT$ deviations in the shape of the landscape. The median difference between $\tilde{\mathcal{F}}_{\mathrm{p}}$ estimated from the two trajectories in this range is $0.3k_BT$, and hence our free-energy analysis gives a consistent result, within SE, between the two different experimental runs.

1. Greenleaf WJ, Woodside MT, Abbondanzieri EA, Block SM (2005) Passive all-optical force clamp for high-resolution laser trapping. *Phys Rev Lett* 95(20):208102.
2. Neuman KC, Block SM (2004) Optical trapping. *Rev Sci Instrum* 75(9):2787–2809.
3. Samuel J, Sinha S (2002) Elasticity of semiflexible polymers. *Phys Rev E Stat Nonlin Soft Matter Phys* 66(5 Pt 1):050801.
4. Kierfeld J, Niamploy O, Sa-yakanit V, Lipowsky R (2004) Stretching of semiflexible polymers with elastic bonds. *Eur Phys J E Soft Matter* 14(1):17–34.
5. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) *Numerical Recipes: The Art of Scientific Computing* (Cambridge Univ Press, Cambridge), 3rd Ed.
6. Hyeon C, Morrison G, Thirumalai D (2008) Force-dependent hopping rates of RNA hairpins can be estimated from accurate measurement of the folding landscapes. *Proc Natl Acad Sci USA* 105(28):9604–9609.
7. Ferrenberg AM, Swendsen RH (1989) Optimized Monte Carlo data analysis. *Phys Rev Lett* 63(12):1195–1198.
8. Shirts MR, Chodera JD (2008) Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys* 129(12):124105.
9. de Messieres M, Brawn-Cinani B, La Porta A (2011) Measuring the folding landscape of a harmonically constrained biopolymer. *Biophys J* 100(11):2736–2744.

10. Gebhardt JCM, Bornschlögl T, Rief M (2010) Full distance-resolved folding energy landscape of one single protein molecule. *Proc Natl Acad Sci USA* 107(5):2013–2018.
11. O'Shea EK, Klemm JD, Kim PS, Alber T (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* 254(5031):539–544.
12. Klimov DK, Betancourt MR, Thirumalai D (1998) Virtual atom representation of hydrogen bonds in minimal off-lattice models of alpha helices: Effect on stability, cooperativity and kinetics. *Fold Des* 3(6):481–496.
13. Denesyuk NA, Thirumalai D (2011) Crowding promotes the switch from hairpin to pseudoknot conformation in human telomerase RNA. *J Am Chem Soc* 133(31):11858–11861.
14. Holtzer ME, et al. (2001) Temperature dependence of the folding and unfolding kinetics of the GCN4 leucine zipper via 13C(alpha)-NMR. *Biophys J* 80(2):939–951.
15. Betancourt MR, Thirumalai D (1999) Pair potentials for protein folding: Choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci* 8(2):361–369.
16. Liu ZX, Reddy G, O'Brien EP, Thirumalai D (2011) Collapse kinetics and chevron plots from simulations of denaturant-dependent folding of globular proteins. *Proc Natl Acad Sci USA* 108(19):7787–7792.
17. Hyeon C, Dima RI, Thirumalai D (2006) Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins. *Structure* 14(11):1633–1645.
18. von Hansen Y, Mehlich A, Pelz B, Rief M, Netz RR (2012) Auto- and cross-power spectral analysis of dual trap optical tweezer experiments using Bayesian inference. *Rev Sci Instrum* 83(9):095116.
19. Peterman EJG, van Dijk MA, Kapitein LC, Schmidt CF (2003) Extending the bandwidth of optical-tweezers interferometry. *Rev Sci Instrum* 74(7):3246–3249.
20. Berg-Sorensen K, Peterman EJG, Weber T, Schmidt CF, Flyvbjerg H (2006) Power spectrum analysis for optical tweezers. II: Laser wavelength dependence of parasitic filtering, and how to achieve high bandwidth. *Rev Sci Instrum* 77(6):063106.
21. Gao Y, Sirinakis G, Zhang Y (2011) Highly anisotropic stability and folding kinetics of a single coiled coil protein under mechanical tension. *J Am Chem Soc* 133(32):12749–12757.
22. Yu H, et al. (2012) Direct observation of multiple misfolding pathways in a single prion protein molecule. *Proc Natl Acad Sci USA* 109(14):5283–5288.
23. Ritchie DB, Foster DAN, Woodside MT (2012) Programmed -1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. *Proc Natl Acad Sci USA* 109(40):16167–16172.
24. Ferdinand AE, Fisher ME (1969) Bounded and inhomogeneous Ising models. I. Specific-heat anomaly of a finite lattice. *Phys Rev* 185(2):832–846.
25. Rabiner LR (1989) A tutorial on hidden Markov-models and selected applications in speech recognition. *Proc IEEE* 77(2):257–286.
26. Baum LE, Petrie T, Soules G, Weiss N (1970) A maximization technique occurring in statistical analysis of probabilistic functions of Markov chains. *Ann Math Stat* 41(1): 164–171.
27. Viterbi AJ (1967) Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans Inf Theory* 13(2):260–269.
28. Miller RG (1974) The jackknife–A review. *Biometrika* 61(1):1–15.
29. Efron B (1979) 1977 Rietz lecture–Bootstrap methods–another look at the jackknife. *Ann Stat* 7(1):1–26.
30. Politis DN, Romano JP (1994) The stationary bootstrap. *J Am Stat Assoc* 89(428): 1303–1313.

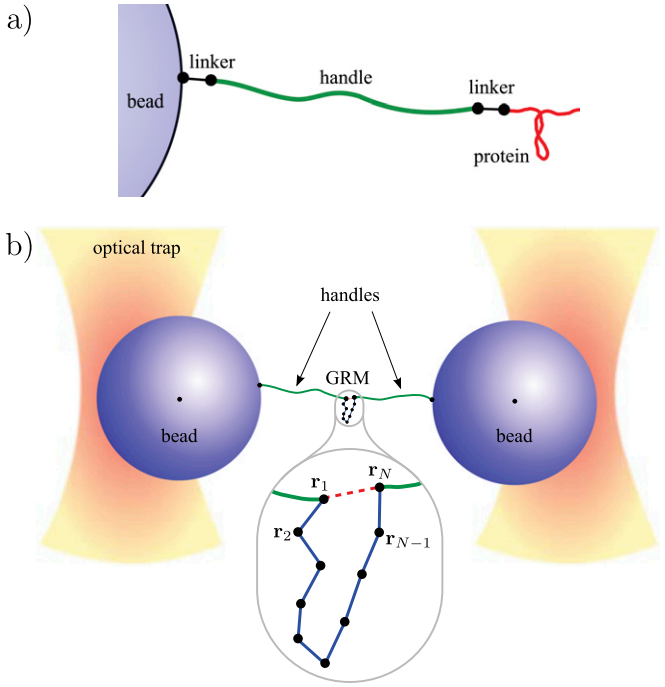

Fig. S1. (*A*) Schematic illustration (not to scale) of the covalent linkers which attach the handle to the bead on one end, and the handle to the protein on the other end. (*B*) GRM within the optical tweezer.

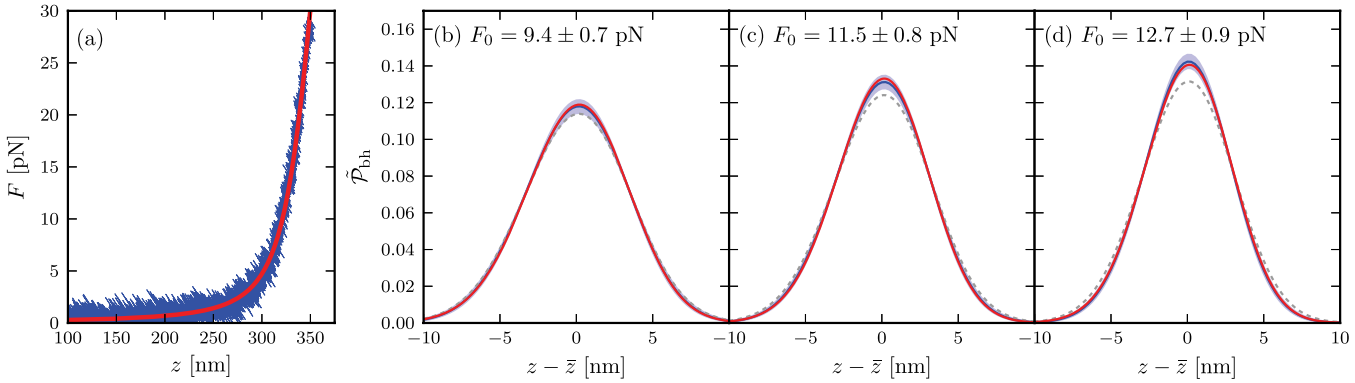

Fig. S2. Experimental (blue solid curves) and theory (red solid curves) results for a system containing only dsDNA handles and beads. The apparatus parameters and theoretical best-fit values are described in *SI Text*, section 2. The same set of best-fit parameters is used for all of the theory curves. (*A*) Force F vs. total extension z from an experimental pulling trajectory, compared with the theoretical mean extension as a function of force. (*B–D*) Total bead–handle probability end-to-end distance distributions $\tilde{\mathcal{P}}_{\rm bh}$ (blue solid curves) derived from three experimental runs at different constant trap separations, corresponding to mean forces of 9.4 ± 0.7, 11.5 ± 0.8, and 12.7 ± 0.9 pN, respectively. In each case the experimental data are corrected for noise/filtering effects using the FBS method (*SI Text*, section 6), and transformed into the constant force ensemble using Eq. **1**, with F_0 chosen to be equal to the mean force value in the trajectory. The distance scale is centered at $\bar{z}$, the mean extension. The light blue shaded region around each blue curve represents the SE margin for every point in the distribution (68% confidence band). For comparison, the experimental results omitting the FBS corrections are shown as gray dashed lines.

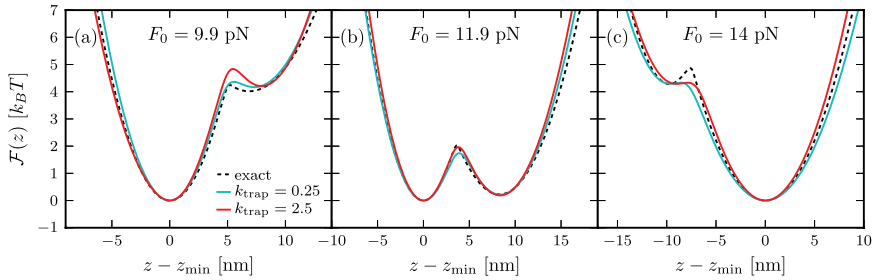

Fig. S3. GRM free energy $\tilde{F}_{\rm GRM}$ after deconvolution for three different values of the force F_0: (*A*) 9.9 pN; (*B*) 11.9 pN; (*C*) 14 pN. In each case, results for two different trap strengths $k_{\rm trap}$ = 0.25, 2.5 pN/nm are shown as solid lines of different color. Exact analytical solutions are drawn as dashed lines. The z scale is plotted relative to $z_{\rm min}$, the location of the minimum in the free energy.

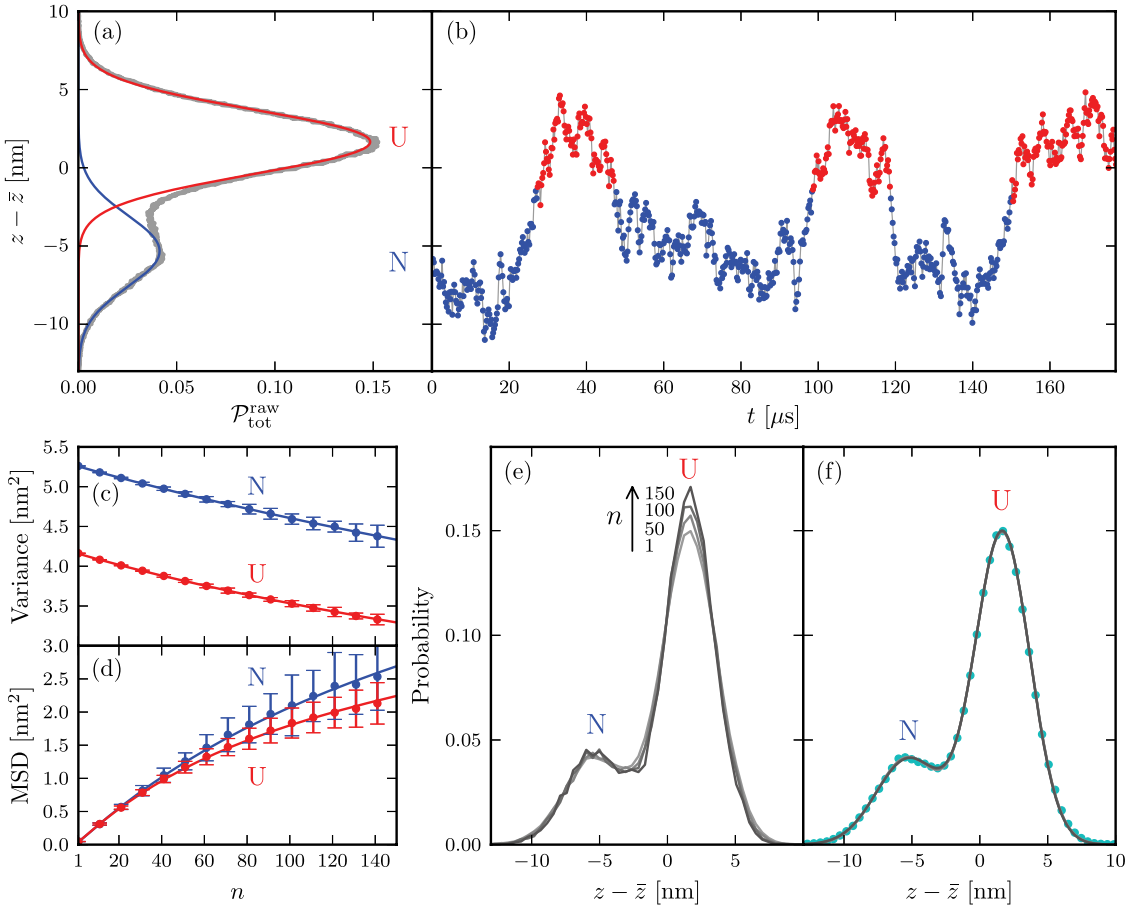

Fig. S4. FBS analysis of a GRM Brownian dynamics simulation ($z_{\rm trap}$ = 1,298 nm; all other parameters as in Table S1). (*A*) Probability distribution of the bead–bead separation from the raw simulation data, $\mathcal{P}_{\rm tot}^{\rm raw}$ (gray), and the decomposition into individual Gaussian peaks corresponding to the N (blue) and U (red) states. Distances are measured with respect to $\bar{z}$, the mean bead–bead separation. (*B*) Sample time series fragment from the simulation, with the individual data points colored according to their assignment to the N (blue) and U (red) states by HMM analysis. (*C*) For the raw time series filtered by averaging together every n data points, the variance $C_{{\rm raw},n}(0)$ as a function of n. The time series corresponding to each state, N and U, is analyzed separately, and plotted as blue and red points respectively, with bars denoting SE due to finite sampling. Solid curves are the FBS theoretical fits to $C_{{\rm raw},n}(0)$ (Eq. **S42**). Best-fit FBS parameters are listed in Table S2. (*D*) Analogous to *C*, except showing the MSD function $\Delta_{{\rm raw},n}(n\tau_s)$ for consecutive pairs of points in the averaged time series. (*E*) Raw distributions $\mathcal{P}_{\rm tot}^{{\rm raw},n}$ for the averaged time series at n = 1, 50, 100, 150. (*F*) Solid curves: distribution $\mathcal{P}_{\rm tot}$ estimated by applying the appropriate FBS correction to the raw distributions in *E*. There are four curves, but due to overlapping they appear as one. Points: raw distribution $\mathcal{P}_{\rm tot}^{{\rm raw},n}$ for n = 1 (no averaging), which for the GRM case is the true distribution, because there are no noise or apparatus filtering effects in the simulation.

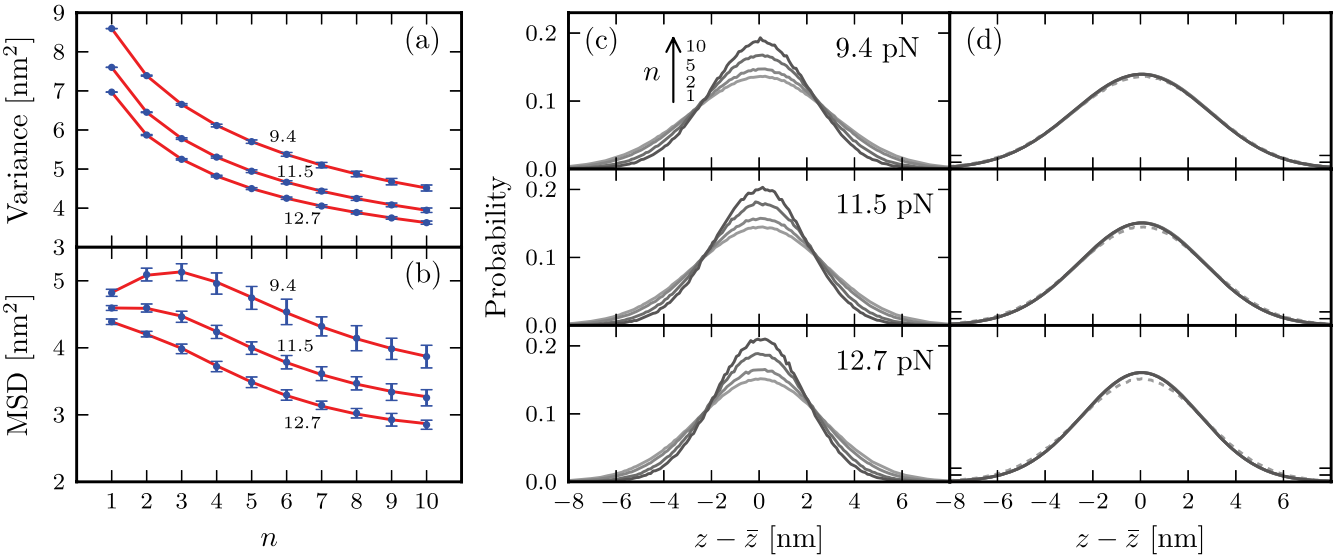

Fig. S5. FBS analysis of experimental results for a system containing only dsDNA handles and beads (see *SI Text*, section 2 for apparatus parameters). We analyze three separate trajectories at different constant trap separations, corresponding to mean forces $F_0 = 9.4 \pm 0.7$, 11.5 ± 0.8, and 12.7 ± 0.9 pN. Results in each panel are labeled by the F_0 value of the trajectory. (*A*) For the raw experimental time series filtered by averaging together every n data points, the variance $C_{\mathrm{raw},n}(0)$ as a function of n. Results are plotted as points, with bars denoting SE due to finite sampling. Solid curves are the FBS theoretical fits to $C_{\mathrm{raw},n}(0)$ (Eq. **S42**). Best-fit FBS parameters are listed in Table S2. (*B*) Analogous to *A*, except showing the MSD function $\Delta_{\mathrm{raw},n}(n\tau_s)$ for consecutive pairs of points in the averaged time series. (*C*) Raw distributions $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},n}$ for the averaged time series at $n = 1, 2, 5, 10$. Each row corresponds to a different trajectory. (*D*) Solid curves: distribution $\mathcal{P}_{\mathrm{tot}}$ estimated by applying the appropriate FBS correction to the raw distributions in *C*. For each trajectory there are four curves, but due to overlapping they appear as one. Dashed curves: raw distribution $\mathcal{P}_{\mathrm{tot}}^{\mathrm{raw},n}$ for $n = 1$. Although this distribution is free of any additional time averaging carried out on the recorded time series, it is subject to parasitic and electronic filtering effects intrinsic to the apparatus. These distortions are corrected by FBS, and hence the dashed and solid curves are distinct.

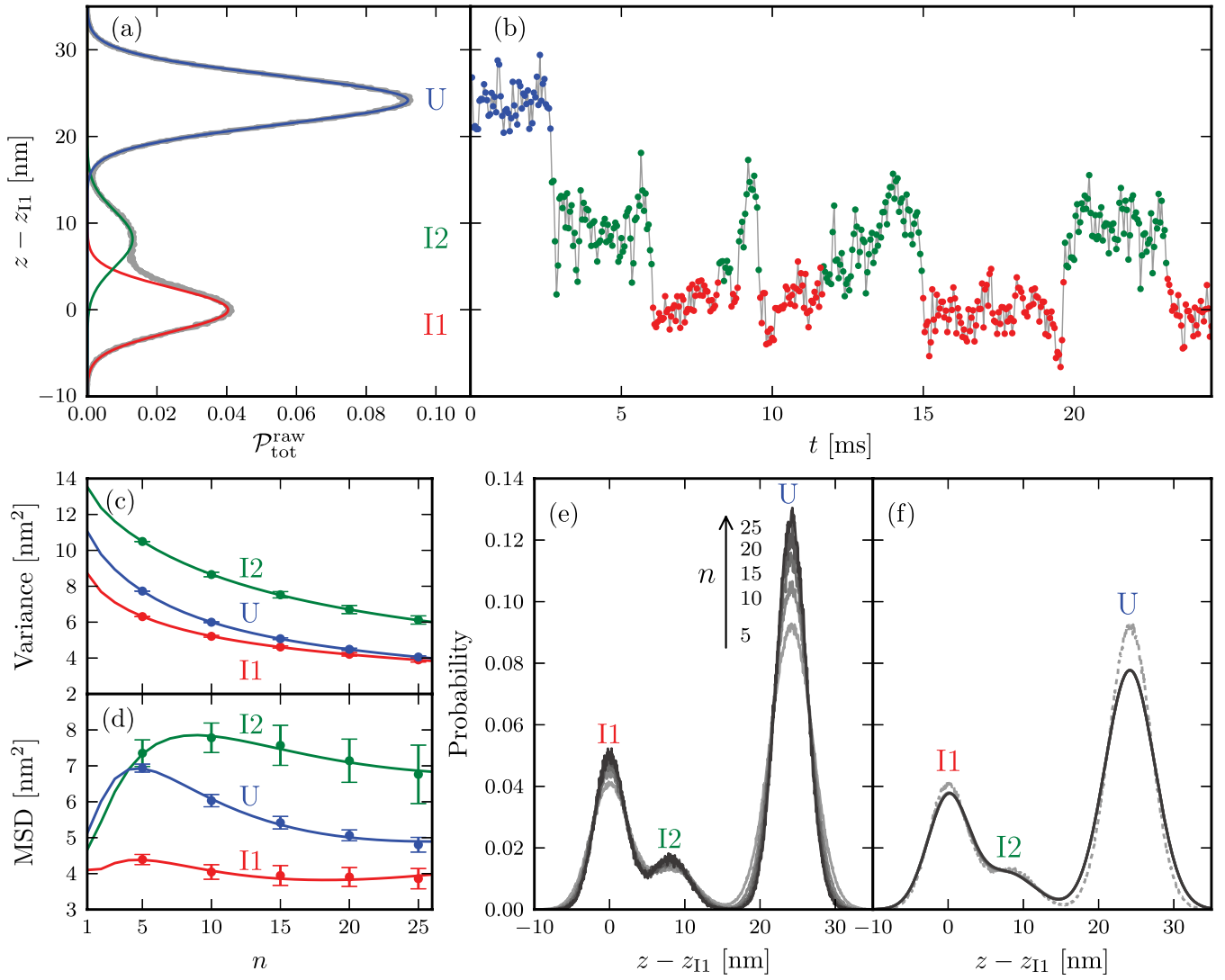

Fig. S6. FBS analysis of a GCN4 leucine zipper experiment (trajectory 1, with parameters given in Table S1). (*A*) Probability distribution of the bead–bead separation from the raw experimental data, $\mathcal{P}_{tot}^{raw}$ (gray), and the decomposition into individual Gaussian peaks corresponding to the I1 (red), I2 (green), and U (blue) states. Distances are measured with respect to z_{I1}, the position of the I1 peak. (*B*) Sample time series fragment from the experiment, with the individual data points colored according to their assignment to the I1 (red), I2 (green), and U (blue) states by HMM analysis. (*C*) For the raw time series filtered by averaging together every n data points, the variance $C_{raw,n}(0)$ as a function of n. Time series corresponding to each state are analyzed separately, and plotted as points in distinct colors, with bars denoting SE due to finite sampling. Solid curves are the FBS theoretical fits to $C_{raw,n}(0)$ (Eq. **S42**). Best-fit FBS parameters are listed in Table S2. (*D*) Analogous to *C*, except showing the MSD function $\Delta_{raw,n}(n\tau_s)$ for consecutive pairs of points in the averaged time series. (*E*) Raw distributions $\mathcal{P}_{tot}^{raw,n}$ for the averaged time series at n = 5, 10, 15, 20, 25. (*F*) Solid curves: the distribution $\mathcal{P}_{tot}$ estimated by applying the appropriate FBS correction to the raw distributions in *E*. There are five curves, but due to overlapping they appear as one. Dashed curve: raw distribution $\mathcal{P}_{tot}^{raw,n}$ for n = 5.

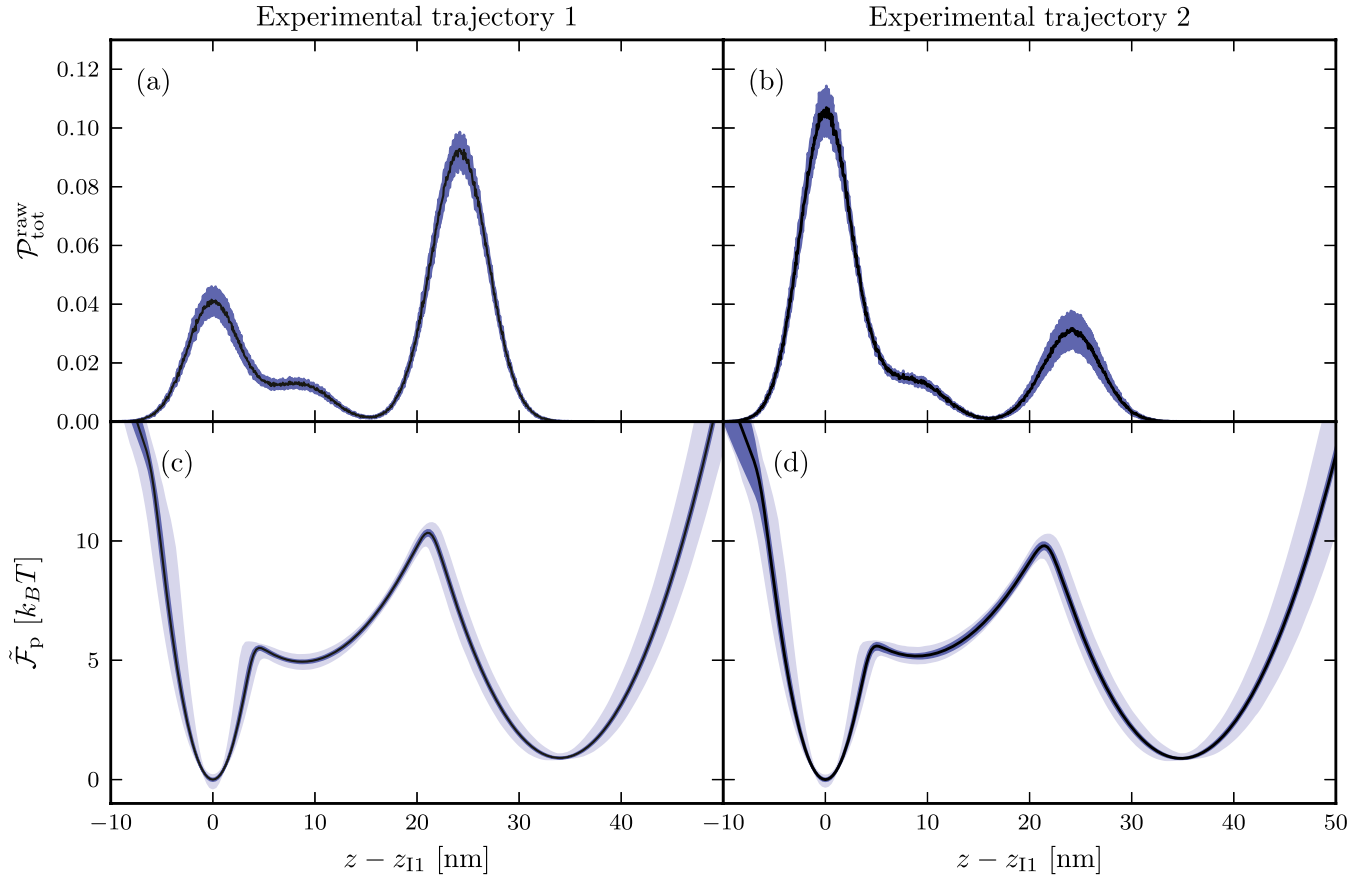

Fig. S7. Estimation of uncertainty in the free-energy reconstruction, as discussed in *SI Text*, sections 7 A and B. (*A* and *B*) Probability distributions $\mathcal{P}_{\rm tot}^{\rm raw}$ of the raw time series for total bead–bead separation (black curves) collected during two experimental runs (left/right columns; see Table S1 for parameters). Distances are measured with respect to $z_{\rm I1}$, the position of the I1 peak. Dark blue band corresponds to the SE (68% confidence interval) for each point in the distribution, due to finite sampling. (*C* and *D*) Corresponding intrinsic protein free energy $\tilde{\mathcal{F}}_{\rm p}$ (black curves), as calculated using the procedure described in the main text. The free energies are in the constant-force ensemble, at the midpoint force value F_0, where the probability of being in states I1 and U is equal (F_0 = 12.3 ± 0.9 pN for run 1; 12.1 ± 0.9 pN for run 2). Dark blue band represents SE (68% confidence interval) including just the uncertainty due to finite sampling; wider light blue band is the SE including all sources of uncertainty (sampling and apparatus parameters).

Table S1. Parameters used in the GRM model and SOP simulation of the LZ26 leucine zipper, together with the corresponding quantities from two experimental runs (10)

Parameter	GRM	SOP simulation	Experiment
Bead: R_b, nm	500	100	500 ± 25
Trap: $z_{\rm trap}$, nm	1,294	483–543	1,553 ± 1, 1,547 ± 1[†]
Trap: $k_{\rm trap}$, pN/nm	0.25	0.25	0.25 ± 0.03, 0.27 ± 0.03[‡]
Trap: α	1/3	1/3	Unknown
Handle: L, nm	100	100	188 ± 2
Handle: l_p, nm	20	20	20 ± 2
Handle: γ, pN	2,780	2,780	400 ± 40
Linker: κ , kcal/mol·nm^2	200	200	200[§]
Linker: ι, nm	1.5	1.5	1.5

[†]Different separations correspond to two folding trajectories.
[‡]Left, right trap strengths.
[§]Linker characteristics are assumed for the experimental case.

Table S2. Parameters used in the FBS analysis of simulation and experimental systems (*SI Text*, section 6)

Case	τ_s (μs)	τ_f (μs)	ν (nm^2 μs)	A_1 (nm^2)	τ_1 (μs)	A_c (nm^2)	B_c (nm^2/μs)
Simulation: GRM							
State *N*	0.024	0	0	2.2 ± 1.2	3.2 ± 1.0	3.1 ± 1.2	0.24 ± 0.15
State *U*	0.024	0	0	1.6 ± 0.2	2.1 ± 0.2	2.6 ± 0.2	0.22 ± 0.04
Experiment: dsDNA handles (no protein)							
F_0 = 9.4 pN	10	7	31.8 ± 1.5	5.0 ± 0.2	14.7 ± 1.8	3.2 ± 0.2	0.0077 ± 0.0020
F_0 = 11.5 pN	10	7	31.8 ± 1.5	4.0 ± 0.2	12.6 ± 1.4	3.0 ± 0.1	0.0079 ± 0.0012
F_0 = 12.7 pN	10	7	31.8 ± 1.5	3.3 ± 0.2	11.5 ± 1.2	2.9 ± 0.1	0.0076 ± 0.0008
Experiment: GCN4 leucine zipper (trajectory 1)							
State I1	10	7	31.8 ± 1.5	3.9 ± 0.2	28.1 ± 5.0	3.5 ± 0.2	0.0066 ± 0.0010
State I2	10	7	31.8 ± 1.5	8.6 ± 0.6	41.9 ± 8.0	4.2 ± 0.8	0.0069 ± 0.0036
State U	10	7	31.8 ± 1.5	7.6 ± 0.1	23.7 ± 1.7	3.2 ± 0.1	0.0065 ± 0.0007
Experiment: GCN4 leucine zipper (trajectory 2)							
State I1	10	7	31.8 ± 1.5	4.1 ± 0.1	28.1 ± 3.5	3.5 ± 0.2	0.0069 ± 0.0008
State I2	10	7	31.8 ± 1.5	8.3 ± 1.0	41.3 ± 12.0	5.3 ± 1.3	0.0074 ± 0.0053
State U	10	7	31.8 ± 1.5	8.1 ± 0.2	23.6 ± 2.2	3.2 ± 0.2	0.0066 ± 0.0009