

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}

^aInstitute for Physical Science and Technology, University of Maryland, College Park, MD 20742; ^bDepartment of Physics, Technical University of Munich, 85748 Garching, Germany; and ^cDepartment 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 $z_{\text{tot}}(t)$ in Fig. 1 as a function of time t]? Here, we solve this problem using a rigorous theoretical procedure. Besides $z_{\text{tot}}(t)$, 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_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 $k_x = k_z \equiv k_{\text{trap}}$ along the lateral plane, and a weaker axial strength $k_y = \alpha k_{\text{trap}}$, 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{z} axis, with trap 1 at $z = 0$ and trap 2 at $z = z_{\text{trap}}$. 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, $\bar{z}_1 = z_{\text{trap}} - \bar{z}_2 = \bar{F}/k_{\text{trap}} > 0$, where \bar{F} is the mean tension along \hat{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 $z_{\text{tot}}(t) \equiv z_2(t) - z_1(t)$ (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.

¹To 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.

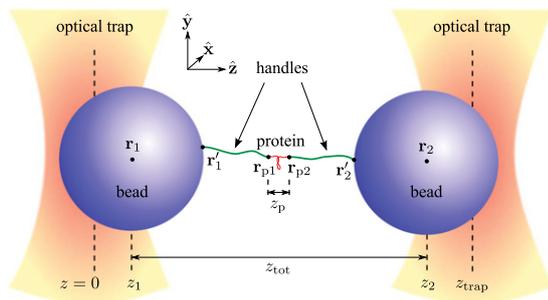


Fig. 1. Dual-beam optical tweezer setup for studying the equilibrium folding landscape of a single protein molecule under force.

approximately constant (in which case we could skip the transformation into the constant-force ensemble described below). The mean tension \bar{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 \bar{F}) or smaller (leading to lower \bar{F}). Because $\bar{F} = k_{\text{trap}}(z_{\text{trap}} - \bar{z}_{\text{tot}})/2$, the precise relationship between z_{trap} and \bar{F} requires knowing the mean total extension \bar{z}_{tot} , which depends among other things on the details of the energy landscape. Hence, we cannot in general calculate beforehand what \bar{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 \bar{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 $\mathcal{P}_{\text{tot}}(z_{\text{tot}})$, 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 $\mathcal{P}_{\text{tot}}(z_{\text{tot}})$, 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 $\mathcal{P}_{\text{tot}}(z_{\text{tot}})$, the objective is to find $\tilde{\mathcal{P}}_{\text{p}}(z_{\text{p}}; F_0)$, 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 $\tilde{\mathcal{F}}_{\text{p}}(z_{\text{p}}; F_0) = -k_B T \ln \tilde{\mathcal{P}}_{\text{p}}(z_{\text{p}}; F_0)$. 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 $\mathcal{P}_{\text{tot}}(z_{\text{tot}})$, we obtain the total system end-to-end distribution at a constant F_0 using

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

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, $\tilde{\mathcal{P}}_{\text{tot}} = \tilde{\mathcal{P}}_{\text{b}} * \tilde{\mathcal{P}}_{\text{h}} * \tilde{\mathcal{P}}_{\text{p}} * \tilde{\mathcal{P}}_{\text{h}} * \tilde{\mathcal{P}}_{\text{b}}$, relates the total end-to-end fluctuations $\tilde{\mathcal{P}}_{\text{tot}}(z_{\text{tot}}; F_0)$ 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{z} . In Fourier space the convolution has the form

$$\begin{aligned} \tilde{\mathcal{P}}_{\text{tot}}(k; F_0) &= \tilde{\mathcal{P}}_{\text{b}}^2(k; F_0) \tilde{\mathcal{P}}_{\text{h}}^2(k; F_0) \tilde{\mathcal{P}}_{\text{p}}(k; F_0) \\ &\equiv \tilde{\mathcal{P}}_{\text{bh}}(k; F_0) \tilde{\mathcal{P}}_{\text{p}}(k; F_0), \end{aligned} \quad [2]$$

where $\tilde{\mathcal{P}}_{\alpha}(k; F_0)$ is the Fourier transform of $\tilde{\mathcal{P}}_{\alpha}(z_{\alpha}; F_0)$. Here, $\tilde{\mathcal{P}}_{\text{bh}}$, 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 $\tilde{\mathcal{P}}_{\text{p}}$ to $\tilde{\mathcal{P}}_{\text{tot}}$. Because $\tilde{\mathcal{P}}_{\text{bh}}$ can be modeled from a theoretical description of the handles and beads, we can solve for $\tilde{\mathcal{P}}_{\text{p}}$ using Eq. 2 and hence find $\tilde{\mathcal{F}}_{\text{p}}$, 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{z}) and transverse (\hat{x}, \hat{y}) components. In general, condition (ii) is not physically sensible (19). However, if $\bar{\rho}_{\text{tot}}$ is the typical length scale describing transverse fluctuations, then condition (i) is approximately valid when $\beta k_{\text{trap}} \bar{\rho}_{\text{tot}}^2 \ll 1$. 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, $\bar{\rho}_{\text{tot}} \sim \mathcal{O}(1 \text{ nm})$, so to ensure condition (i) is met, we require that $k_{\text{trap}} \ll k_B T / \bar{\rho}_{\text{tot}}^2 = 4.1 \text{ pN/nm}$ at $T = 298 \text{ K}$. We use the experimental value $k_{\text{trap}} = 0.25 \text{ pN/nm}$ 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 $F_0 = \bar{F}$, the mean tension, is a reasonable choice.

Step *ii* depends on knowledge of $\tilde{\mathcal{P}}_{\text{bh}}(k; F_0)$, 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 $\tilde{\mathcal{P}}_{\text{h}}$ 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 $\tilde{\mathcal{P}}_{\text{bh}}(k; F_0)$. 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 $\mathcal{P}_{\text{tot}} = \mathcal{P}_{\text{bh}}$ (*SI Text*). The theory simultaneously fits results for several experimental quantities measured on the same system: the distributions $\tilde{\mathcal{P}}_{\text{bh}}$ derived from three different trap separations, corresponding to mean forces $F_0 = 9.4\text{--}12.7 \text{ pN}$, and a force-extension curve. The accuracy of the model $\tilde{\mathcal{P}}_{\text{bh}}$ is $\sim 1\text{--}3\%$, 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 $\tilde{\mathcal{P}}_{\text{GRM}}$ at $F_0 = 11.9 \text{ pN}$ is plotted in Fig. 2A. The upper part shows a projection onto the $(\rho = \sqrt{x^2 + y^2}, z)$ plane, because $\tilde{\mathcal{P}}_{\text{GRM}}$ 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 $\tilde{\mathcal{P}}_h$ and $\tilde{\mathcal{P}}_b$ (including transverse fluctuations) are illustrated in Fig. 2B and C. The full 3D behavior is derived in an analogous manner to the theory mentioned above for the 1D Fourier-space distribution $\tilde{\mathcal{P}}_{bh}(k; F_0)$ 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}_{tot} in Fig. 2D. The bead, handle, linker, and trap parameters are listed in Table S1. From \mathcal{P}_{tot} one can calculate the mean total z extension and the mean tension, which in this case are $\bar{z}_{tot} = 1,199$ nm, $\bar{F} = k_{trap}(z_{trap} - \bar{z}_{tot})/2 = 11.9$ pN.

The \bar{z} -probability projection in Fig. 2D (Lower) is the information accessible in an experiment, and the computation of the intrinsic distribution in Fig. 2A (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. 2E–G illustrates the construction procedure for the GRM optical tweezer system. E corresponds to step i , with a transformation of the distribution \mathcal{P}_{tot} (whose varying force scale is shown along the top axis) into $\tilde{\mathcal{P}}_{tot}$ at constant force $F_0 = 11.9$ pN. Step ii uses the exact $\tilde{\mathcal{P}}_{bh}$, shown in real space in F, and produces the intrinsic distribution $\tilde{\mathcal{P}}_{GRM}$, 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. 3A we compare the free energy $\mathcal{F}_{tot} = -k_B T \ln \mathcal{P}_{tot}$ for a fixed $L = 100$ nm and a varying l_p/L , derived from the simulation trajectories, and the exact intrinsic GRM result $\tilde{\mathcal{F}}_{GRM} = -k_B T \ln \tilde{\mathcal{P}}_{GRM}$ 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}_{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. 3B). For each \mathcal{F}_{tot} in Fig. 3A, Fig. 3B compares the result of the construction procedure and the exact answer for $\tilde{\mathcal{F}}_{GRM}$. 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 $\tilde{\mathcal{F}}_{GRM}$ 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_{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 \bar{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}_{tot} derived by convolving the component probabilities and accounting for the optical traps. (E–G) Construction of the original GRM distribution $\tilde{\mathcal{P}}_{GRM}$ starting from \mathcal{P}_{tot} . (E) \mathcal{P}_{tot} (purple) and $\tilde{\mathcal{P}}_{tot}$ (blue) as a function of z on the bottom axis, measured relative to \bar{z} , the average extension for each distribution. For \mathcal{P}_{tot} , the upper axis shows the z range translated into the corresponding trap forces F . After removing the trap effects, $\tilde{\mathcal{P}}_{tot}$ is the distribution for constant force $F_0 = 11.9$ pN. (F) $\tilde{\mathcal{P}}_{bh}$, describing the total probability at F_0 of fluctuations resulting from both handles and the rotation of the beads. (G) Constructed solution for $\tilde{\mathcal{P}}_{GRM}$ (solid line), obtained by numerically inverting the convolution $\tilde{\mathcal{P}}_{tot} = \tilde{\mathcal{P}}_{bh} * \tilde{\mathcal{P}}_{GRM}$. Exact analytical result for $\tilde{\mathcal{P}}_{GRM}$ is shown as a dashed line; z_N is the position of the N peak.

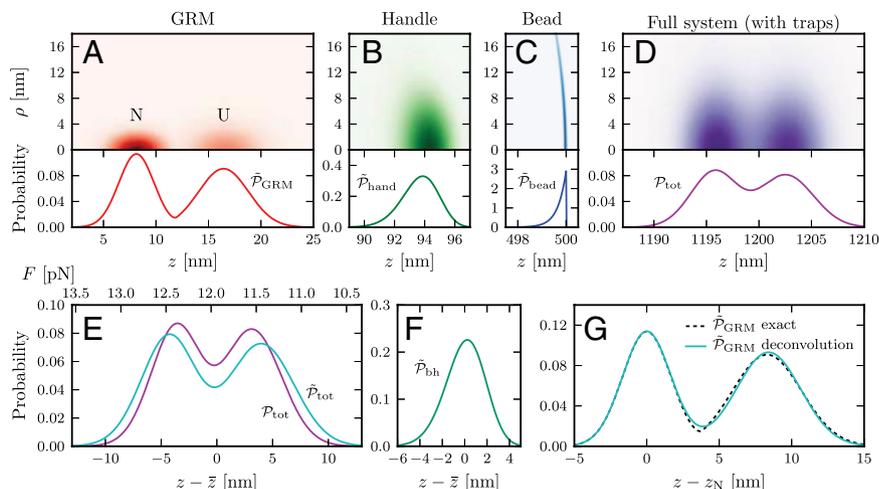
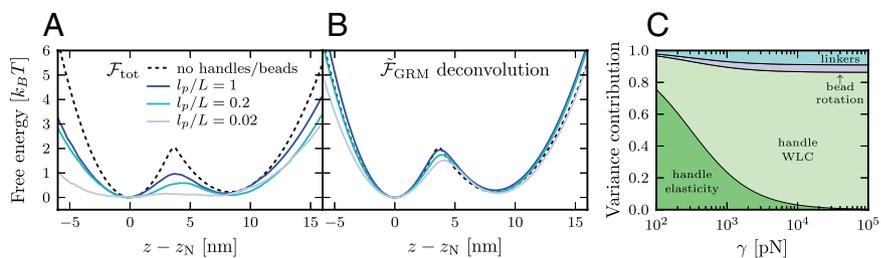


Fig. 3. Effects of handle characteristics on the free-energy profile of the GRM in a LOT setup. (A) Total system free energy $\mathcal{F}_{\text{tot}} = -k_B T \ln \mathcal{P}_{\text{tot}}$ for fixed $L = 100$ nm, and varying ratios l_p/L . All of the other parameters are in Table S1 (GRM column). The exact analytical free energy at $F_0 = 11.9$ pN (dashed line) for the GRM alone, $\tilde{\mathcal{F}}_{\text{GRM}} = -k_B T \ln \tilde{\mathcal{P}}_{\text{GRM}}$, is shown for comparison. (B) For each \mathcal{F}_{tot} in A, the construction of $\tilde{\mathcal{F}}_{\text{GRM}}$ 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 $\tilde{\mathcal{P}}_{\text{bh}}$ 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 $\tilde{\mathcal{F}}_p = -k_B T \ln \tilde{\mathcal{P}}_p$ at $F_0 = 12.3$ pN in Fig. 4A has four prominent wells in $\tilde{\mathcal{F}}_p$ as a function of z_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 $k_B T$, 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. 5A) at constant force $F_0 = 12.3$ pN, and within the full optical tweezer system (Fig. 5C) with $z_{\text{trap}} = 503$ nm. For the latter case we plot both $z_{\text{tot}}(t)$ (purple) and $z_p(t)$ (gray), allowing us to see how the bead separation tracks changes in the protein extension. The probability distributions $\tilde{\mathcal{P}}_p$ and \mathcal{P}_{tot} are plotted in Fig. 5B and D, respectively. In Fig. 5E, the distribution \mathcal{P}_{tot} within the optical tweezer system is plotted for $z_{\text{trap}} = 503$ nm. Although we only illustrate this particular z_{trap} value, ~ 260 trajectories are generated at different z_{trap} and combined together using WHAM (22) (*SI Text*) to produce a single $\tilde{\mathcal{P}}_{\text{tot}}$ at a constant force $F_0 = 12.3$ pN (Fig. 5E). We can then use our theoretical method to recover the protein free energy $\tilde{\mathcal{F}}_p$ (Fig. 5F). 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 $\tilde{\mathcal{F}}_p$ derived from protein-only simulations.

This is particularly striking given that the total system free energy $\mathcal{F}_{\text{tot}}(z_{\text{tot}}) = -k_B T \ln \mathcal{P}_{\text{tot}}(z_{\text{tot}})$ (Fig. 5F), 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 $\mathcal{F}_{\text{tot}}(z_{\text{tot}})$ 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 $\tilde{\mathcal{F}}_p$. 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 $\tilde{\mathcal{F}}_p$. 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_{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 $\tilde{\mathcal{P}}_p$. 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 $\tilde{\mathcal{F}}_p$ is shown for both runs in Fig. 5H (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 $0.4 k_B T$ in the z range shown (see *SI Text* for full error analysis). The landscapes from the two independent runs have a median difference of $0.3 k_B T$, 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 $\tilde{\mathcal{F}}_p$ is a testament to the stability of the dual optical tweezer setup. Each trajectory lasted for more than 100 s, and thus collected $\sim (10^2 - 10^5)$ of the various types of transitions between protein states (the slowest transition, $U \rightarrow I2$, occurred on time scales of 0.4–0.6 s).

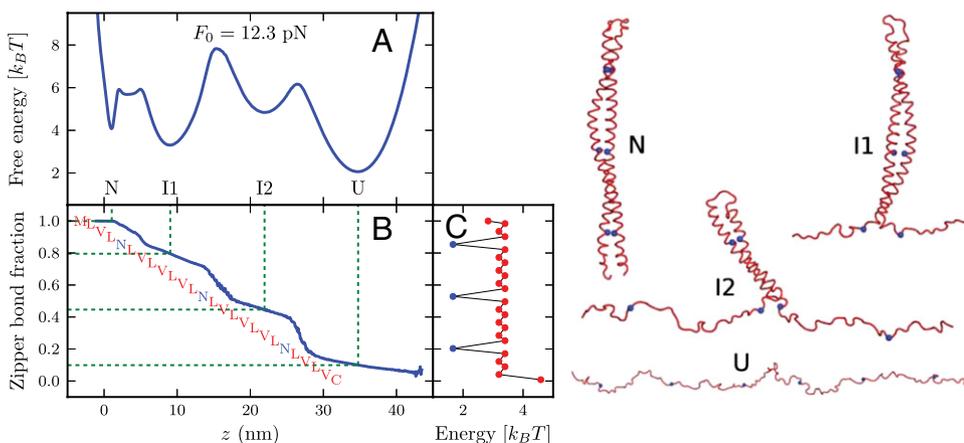


Fig. 4. Intrinsic characteristics of the LZ26 leucine zipper at constant F_0 , derived from SOP simulations in the absence of handles/beads. (A) LZ26 free energy $\tilde{\mathcal{F}}_p$ at $F_0 = 12.3$ pN vs. end-to-end extension z . (Right) Representative protein configurations from the four wells (N, 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].

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, $\tau_r \sim 20 \mu\text{s}$ 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 \rightarrow N \rightarrow 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 $z_{\text{tot}}(t)$ and $z_p(t)$, compared in two different trajectory fragments in Fig. 5C. 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 $t = 36\text{--}42 \mu\text{s}$. When the protein makes sharp, extremely brief excursions (such as a visit to the N state from I1 in the enlarged interval $t = 90\text{--}96 \mu\text{s}$), 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}_{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.

- Thirumalai D, Hyeon C (2005) RNA and protein folding: Common themes and variations. *Biochemistry* 44(13):4957–4970.
- Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: The energy landscape perspective. *Annu Rev Phys Chem* 48:545–600.
- Dill KA, Ozkan SB, Shell MS, Weikl TR (2008) The protein folding problem. *Annu Rev Biophys* 37:289–316.
- Thirumalai D, O'Brien EP, Morrison G, Hyeon C (2010) Theoretical perspectives on protein folding. *Annu Rev Biophys* 39:159–183.
- Guo Z, Thirumalai D (1995) Kinetics of protein folding: Nucleation mechanism, time scales, and pathways. *Biopolymers* 36(1):83–102.
- Klimov DK, Thirumalai D (2005) Symmetric connectivity of secondary structure elements enhances the diversity of folding pathways. *J Mol Biol* 353(5):1171–1186.
- Stigler J, Ziegler F, Gieseke A, Gebhardt JCM, Rief M (2011) The complex folding network of single calmodulin molecules. *Science* 334(6055):512–516.
- Woodside MT, et al. (2006) Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* 314(5801):1001–1004.
- 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.
- 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.
- Gebhardt JCM, Bornschlöggl T, Rief M (2010) Full distance-resolved folding energy landscape of one single protein molecule. *Proc Natl Acad Sci USA* 107(5):2013–2018.
- 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.
- Hyeon C, Thirumalai D (2006) Forced-unfolding and force-quench refolding of RNA hairpins. *Biophys J* 90(10):3410–3427.
- Manosas M, et al. (2007) Force unfolding kinetics of RNA using optical tweezers. II. Modeling experiments. *Biophys J* 92(9):3010–3021.
- 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.
- 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.
- Hummer G, Szabo A (2010) Free energy profiles from single-molecule pulling experiments. *Proc Natl Acad Sci USA* 107(50):21441–21446.
- Hinczewski M, von Hansen Y, Netz RR (2010) Deconvolution of dynamic mechanical networks. *Proc Natl Acad Sci USA* 107(50):21493–21498.
- 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.
- 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.
- Neuman KC, Block SM (2004) Optical trapping. *Rev Sci Instrum* 75(9):2787–2809.
- Ferrenberg AM, Swendsen RH (1989) Optimized Monte Carlo data analysis. *Phys Rev Lett* 63(12):1195–1198.
- Shirts MR, Chodera JD (2008) Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys* 129(12):124105.
- 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.
- Kierfeld J, Niamply O, Sa-yakanit V, Lipowsky R (2004) Stretching of semiflexible polymers with elastic bonds. *Eur Phys J E Soft Matter* 14(1):17–34.
- Bornschlöggl T, Rief M (2006) Single molecule unzipping of coiled coils: Sequence resolved stability profiles. *Phys Rev Lett* 96(11):118102.
- 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.
- Bornschlöggl T, Rief M (2008) Single-molecule dynamics of mechanical coiled-coil unzipping. *Langmuir* 24(4):1338–1342.
- 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.
- 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.
- 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.
- 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.
- Ermak DL, McCammon JA (1978) Brownian dynamics with hydrodynamic interactions. *J Chem Phys* 69(4):1352–1360.