Supporting Information

Single-molecule observation of helix staggering, sliding, and coiled coil misfolding

Zhiqun Xi, Ying Gao, George Sirinakis, Honglian Guo, and Yongli Zhang

Department of Cell Biology, Yale University School of Medicine, 333 Cedar St., New Haven, CT 06520, USA.

Correspondence should be addressed to Y.Z. (yongli.zhang@yale.edu)

Supporting information includes text, one table, and nine figures.

SI Text

Data analysis

Force, extension, contour length, and trap separation

Figure S1 shows our experimental setup to study folding of a single protein using dual-trap optical tweezers. We directly control the separation between the two optical traps to regulate the mechanical tension exerted on the protein. At a fixed trap separation (D), the protein may spontaneous unfold and refold due to thermal fluctuations, leading to the corresponding tether extension (x) and tension (F) changes that can be detected by the optical tweezers. Thus different folding states can be characterized by the contour length (l) or number of amino acids of the unfolded polypeptide under tension. Therefore, we choose the contour length or the number of the unfolded amino acids as the reaction coordinate to describe the folding reaction, adopting a value of 0.4 nm for the contour length per amino acid (1). The contour length is a better choice for the reaction coordinate than the protein extension widely used in the case of constant force. The contour length is directly related to the protein structure and decreases as protein folds, whereas the extension is not well defined in the variable force case and may increase during folding. For example, a single-stranded DNA molecule increases its extension when hybridized with its complementary strand under a tension less than about 6 pN (2, 3). Importantly, the contour length can also serve as a reaction coordinate for protein folding in the absence of an external force (**Fig. S9**), whereas the extension of any unfolded polypeptide is always zero and cannot be defined as a reaction coordinate. Thus, we will express force, extension, and energy of the system in terms of the contour length. First we can write down the following equation based upon **Fig. S1**:

$$\frac{F}{k_{trap}} + x_{DNA} + x_m + h + R_1 + R_2 = D. \qquad \textbf{[S1]}$$

The force-extension curves (FEC) of the DNA handle and the unfolded polypeptide are described by the Marko-Siggia formulas (4, 5), i.e.,

$$F = \frac{k_B T}{P_{DNA}} \left[\frac{1}{4\left(1 - \frac{x_{DNA}}{L_{DNA}}\right)^2} + \frac{x_{DNA}}{L_{DNA}} - \frac{1}{4} \right], \qquad \textbf{[S2]}$$

and

$$F = \frac{k_B T}{P_m}\left[\frac{1}{4\left(1-\frac{x_m}{l}\right)^2}+\frac{x_m}{l}-\frac{1}{4}\right] \quad \textbf{[S3]}$$

, respectively. In our analysis, we adopted the values for DNA persistence length P_{DNA} =50 nm and contour length L_{DNA} =768 nm and polypeptide persistence length P_{DNA} =0.6 nm and the contour length l ranging from 3.2 nm for the fully folded coiled coil and 32 nm for the fully unfolded coiled coil. Equations [S2] and [S3] define extensions of the DNA and the polypeptide as a function of force F and can be substituted into Eq. [S1] to numerically solve for the force and the extensions, given the reaction coordinate l.

Free energy of the dumbbell system

The total free energy of the dumbbell system shown in Fig. S1 includes the potential energy of the two trapped beads ($F^2/(2k_{trapp})$), the entropic energy of the DNA handle (E_{DNA}), the entropic energy of the unfolded polypeptide (E_m) and the protein folding energy (G), i.e.,

$$E(D,l) = \frac{F^2}{2k_{trap}} + E_{DNA}(F) + E_m(l) + G(l) \quad \textbf{[S4]}$$

, where the entropic energy of a worm-like chain under a tension F can be calculated by integration of the Marko-Siggia formula over extension (6), i.e.,

$$E_{DNA}(F) = \frac{k_B T}{P_{DNA}}\frac{L_{DNA}}{4\left(1-\frac{x_{DNA}}{L_{DNA}}\right)}\left[3\left(\frac{x_{DNA}}{L_{DNA}}\right)^2 - 2\left(\frac{x_{DNA}}{L_{DNA}}\right)^3\right], \quad \textbf{[S5]}$$

and

$$E_m(l) = \frac{k_B T}{P_m}\frac{l}{4\left(1-\frac{x_m}{l}\right)}\left[3\left(\frac{x_m}{l}\right)^2 - 2\left(\frac{x_m}{l}\right)^3\right]. \quad \textbf{[S6]}$$

The folding energy $G(l)$ defines the energy landscape of protein folding/unfolding as a function of the contour length of the unfolded polypeptide l at zero force. Because the equilibrium force F and the extensions x_{DNA} and x_m can be calculated from the length constraints in Eqs. [S1], [S2], and [S3], Eq. [S4] defines the energy landscape of the protein folding and unfolding stretched by the dual trap optical tweezers, given the energy landscape $G(l)$ at zero force.

Thus, the probability density of the protein being in state l at any given trap separation D can be calculated using the standard statistical method, i.e.,

$$\rho(D,l)=\frac{\exp\left[-\frac{E(D,l)}{k_BT}\right]}{\int\exp\left[-\frac{E(D,l')}{k_BT}\right]d\Omega}, \qquad \textbf{[S7]}$$

where the integral in the denominator integrates the whole phase space.

State energy and transition rates

The energy landscapes $G(l)$ and $E(D,l)$ only depend on the contour length, but are defined in the complex phase space of protein folding. This can be understood in analog with the gravitational potential energy that is defined in 3D space but only depends on one parameter height. Thus it is generally impossible to deconvolute the multi-dimensional energy landscape from the one-dimensional distance measurements. Fortunately, the energy and kinetics of protein folding can be well described by some minimal and saddle points in the multidimensional phase space that define the different folding states, transition states, and the transition pathways between these states (for a 1D case, see **Fig. S9**). The stable or metstable folding states can be designed by their contour length $l_i, i=1,\cdots,n$, where n is the total number of states that can be determined from the single-molecule experiments. Similarly, the transition states between two different states, i and j, can be designed by the contour length $l_{ij}, i<j$. Finally, the free energy of these folding states and transition states *at zero force* can be designed as G_i and $G_{ij}, i<j$, respectively. Then we can calculate the corresponding state energy at trap separation D,

$$E_i(D)\equiv E(D,l_i)=\frac{F_i^2}{2k_{trap}}+E_{DNA}(F_i)+E_m(l_i)+G_i, \qquad \textbf{[S8]}$$

and the transition energy between states i and j

$$E_{ij}(D)\equiv E(D,l_{ij})=\frac{F_{ij}^{\ 2}}{2k_{trap}}+E_{DNA}(F_{ij})+E_m(l_{ij})+G_{ij}, i<j \qquad \textbf{[S9]}$$

, where the average force F_i and F_{ij} and the associated entropic energy are calculated based on specific contour lengths l_i and l_{ij}, respectively, using Eqs. [S1]-[S6]. Therefore, the probability of the protein being in folding state i can be calculated as a function of trap separation

$$\rho_i(D)=\frac{\exp\left[-\frac{E_i(D)}{k_BT}\right]}{\sum_{j=1}^{n}\exp\left[-\frac{E_j(D)}{k_BT}\right]}. \quad \textbf{[S10]}$$

The nominal activation energy for the protein to transit from state i to state j can be calculated using Eqs. [S8] and [S9], i.e.,

$$\Delta G_{ij}^{\dagger}(D)=E_{ij}(D)-E_i(D). \quad \textbf{[S11]}$$

Then the transition rate from state i to state j is

$$k_{ij}(D)=\begin{cases}k_m\exp\left(-\Delta G_{ij}^{\dagger}(D)/k_BT\right), & \Delta G_{ij}^{\dagger}(D)>0;\\ k_m, & \Delta G_{ij}^{\dagger}(D)\le 0.\end{cases} \quad \textbf{[S12]}$$

Here k_m =1×10^6 s^{-1} is the molecular transition rate (7). The positive energy difference represents a true energy barrier (8), whereas the negative energy difference indicates down-hill protein folding and may occur at zero force (**Fig. S9B**). Note that the opposite transition from state j to state i passes through the same nominal transition state with contour length l_{ij} and energy $E_{ij}(D)$, the nominal activation energy and transition rate can be written as

$$\Delta G_{ji}^{\dagger}(D)=E_{ij}(D)-E_j(D) \quad \textbf{[S13]}$$

and

$$k_{ji}(D)=\begin{cases}k_m\exp\left(-\Delta G_{ji}^{\dagger}(D)/k_BT\right), & \Delta G_{ji}^{\dagger}(D)>0,\\ k_m, & \Delta G_{ji}^{\dagger}(D)\le 0,\end{cases} \quad \textbf{[S14]}$$

respectively.

In conclusion, given the structural and energy parameters for different folding and transition states ($l_i, l_{ij}, G_i, G_{ij}, i<j, i, j=1,\cdots,n$) and experimental parameters ($k_{trap}, L_{DNA}, P_{DNA}, P_m$), one can calculate the state probability and the transition rates between these states and compare them with the experimental data.

Application to the coiled coil

We have previously tested our above theory on the GCN4 coiled and obtained folding energy and transition kinetics consistent with the corresponding ensemble measurements (6). To determine the parameters for pIL folding and misfolding, we first calculated the contour lengths for the fully folded and three partially folded states as illustrated in **Fig. S3**. The unfolded state has a contour length $l_5 = 32$ nm for the crosslinked polypeptide with 80 amino acids. The extension of the structural portion of the partially and fully folded coiled coil, including their associated transition states, is chosen as 3 nm, or $h = 3$ nm in Eq. [S1]. For the unfolded state, $h = 0$ nm. In addition, we choose the unfolded state as our energy reference, i.e., zero. A total of 16 parameters in three sets, the folding energy, the transition state energy, and the location in terms of the contour length, for the four partially and fully folded states are determined by simultaneously fitting the theoretical predictions against the measured state forces, probabilities, and the transition rates. The nonlinear least-square data fitting was computed using a MATLAB program previously used (6). The average of the best-fit parameters are then corrected for the fraying ends by linearly scaling contour lengths between that of the fully folded state l_1=8 amino acids and that of the unfolded state $l_5 = 80$ amino acids. The final values are listed in **Fig. 3** and in **Table S1**.

SI Table

Table S1. The free energy and contour length of different folding states and their corresponding transition states derived from the single-molecule experiments. All parameters here correspond to zero force case and are the averages of the best-fit parameters from experiments on 14 different pIL molecules. Since the free energy of the unfolded state is chosen as zero, a negative transition state energy indicates a down-hill protein folding process. A barrier-limited protein folding process in the presence of force or denaturant may be barrier-less at zero force or in the absence of the denaturant (**Fig. S9B**) (9-11).

State #	Folding energy (k_BT)	State position (a.a.)	Transition state energy (k_BT)	Barrier location (a.a.)	Folding rate $\log_{10} k\left(s^{-1}\right)$
1	-24 (1)	8	-2 (2)	31 (2)	6.0 (0.8)
2	-17 (2)	15 (1)	-2 (2)	28 (3)	6.0 (0.8)
3	-13 (2)	22 (1)	1 (3)	34 (4)	5.5 (1.3)
4	-9 (3)	30 (3)	11 (3)	41 (4)	1.4 (1.3)
5	0	80	N.A.	N.A.	N.A.

SI Figures

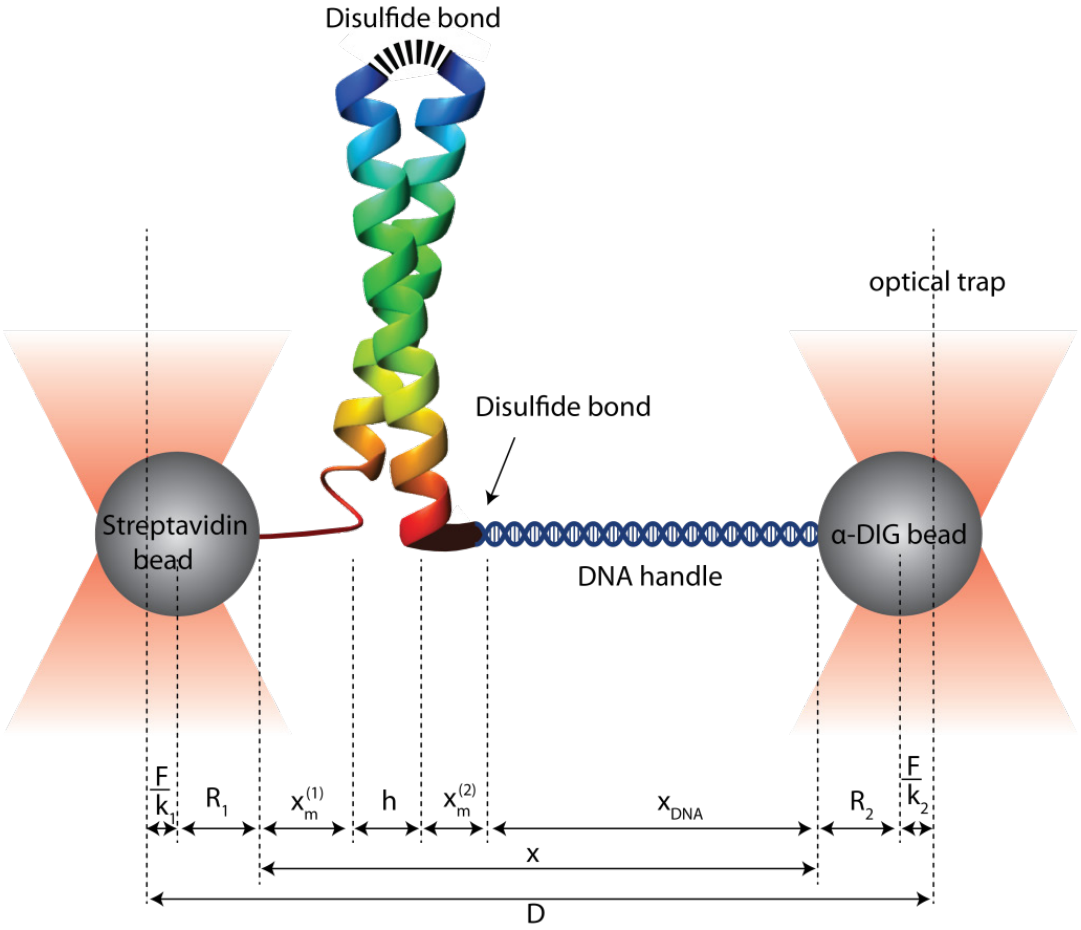

Fig. S1 Experimental setup to pull a single coiled coil protein using dual-trap optical tweezers. The coiled coil is attached to the streptavidin-coated polystyrene bead via a biotin moiety and to the anti-digoxigenin-coated bead via a 2,260 bp DNA handle. The protein and the DNA handle are joined by a disulfide bond under oxidized condition. Some length parameters relevant to the method of data analysis in **SI Text** are indicated, including trap separation D, bead radius R_1 and R_2, and DNA handle extension x_{DNA}. The protein will change its extension upon folding and unfolding. The total extension of the protein is comprised by the extensions of the unfolded polypeptide portion ($x_m = x_m^{(1)} + x_m^{(2)}$) and the folded structural portion or the diameter of the coiled coil here (h =3 nm) (12). The total displacement of the two beads is

$$x_{bead} = \frac{F}{k_1} + \frac{F}{k_2} = \frac{F}{k_{trap}}, \qquad \textbf{[S15]}$$

where F is the tension in the protein-DNA tether and k_1, k_2, and k_{trap} are the stiffness of the individual traps and their combination in series, or

$$k_{trap} = \frac{k_1 k_2}{k_1 + k_2}. \qquad \textbf{[S16]}$$

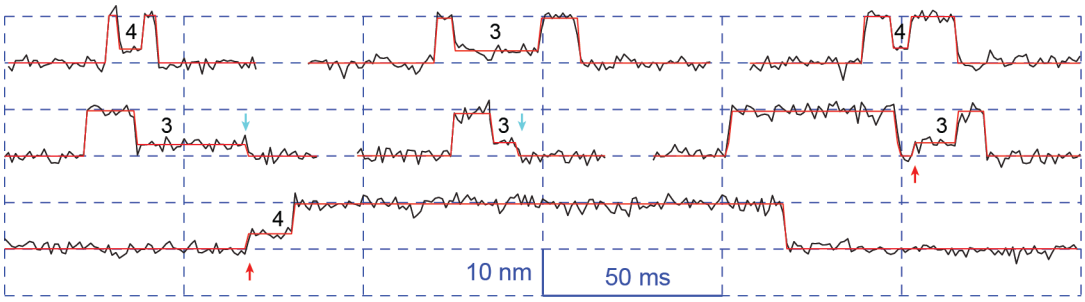

Fig. S2 Partially folded states (or staggered states) are observed as misfolded states (in traces in the first row) or intermediate states (the second and third rows). The red lines are the corresponding idealized HMM fit. The intermediates are formed by helix sliding and can be intermediates for pIL folding (marked by cyan arrows) or unfolding (red arrows). Each state is labeled with the corresponding state number based on its average relative extension (**Fig. 1E** and **Fig. 3**).

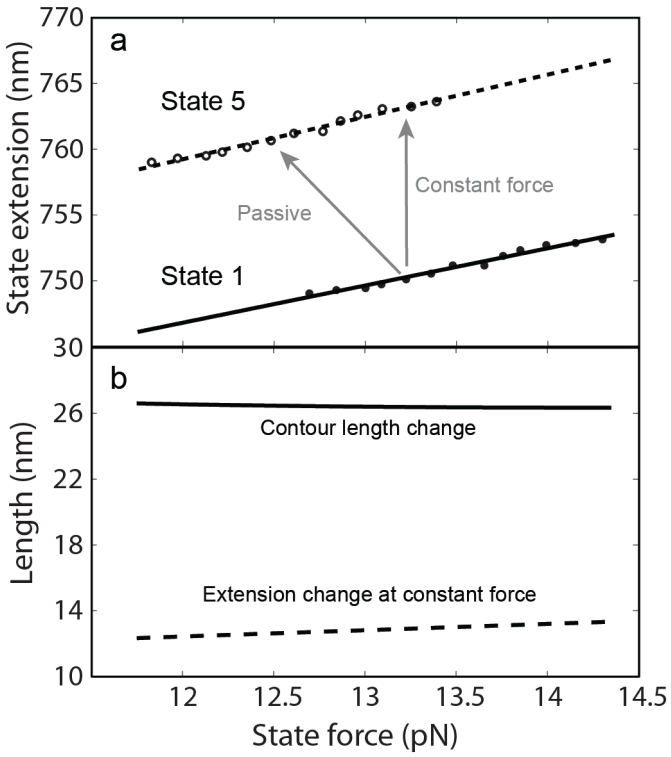

Fig. S3 Change in the contour length of the unfolded polypeptide can be determined by the measured state extension change. (**a**) Force-dependent extensions of the folded and unfolded states (symbols) and their linear fit (lines). Throughout this work, the extension is defined as the end-to-end DNA-protein tether length, which is the measured distance between centers of the beads subtracted by the sum of their radii (6). Note that our optical tweezers were operated in a passive mode, in which protein unfolding is accompanied by a force decrease (indicated by the arrow). As a result, the observed extension change depends not only on the contour length change and persistence length (0.6 nm) of the polypeptide unfolded, but also on the compliances of optical traps and DNA handle. To remove the contribution from the latter, we linearly fit the two state positions and calculated the corresponding extension changes at the constant force. (**b**) The force-dependent extension change at a constant force (dashed line) and the corresponding contour length change derived from the worm-like chain model (4) (solid line). The contour length change does barely vary with the force, indicating a relatively rigid folded state in the tested force range.

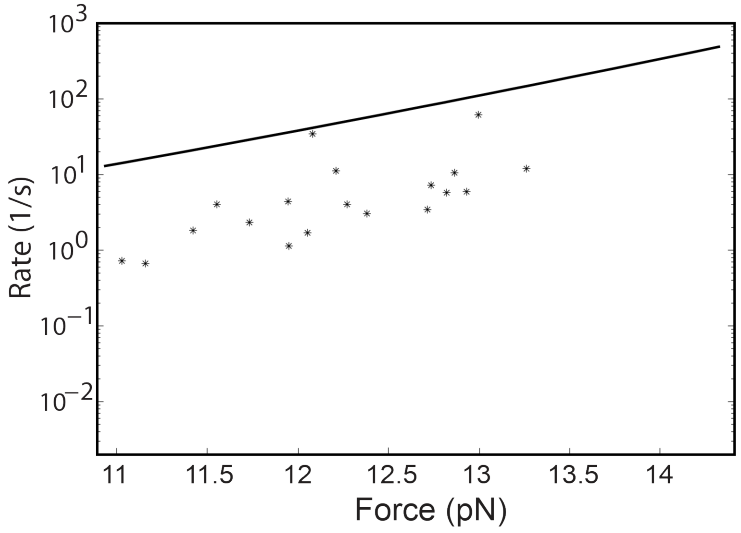

Fig. S4 The measured (stars) and predicted (line) unfolding rates of pIL from the partially and misfolded state with four heptad shifts (or State 4 illustrated in **Fig. 3**). Due to its low population, this state has larger error in its measured unfolding rates than other states, which is indicated by the relatively larger standard deviation of the derived transition energy (**Table S1**).

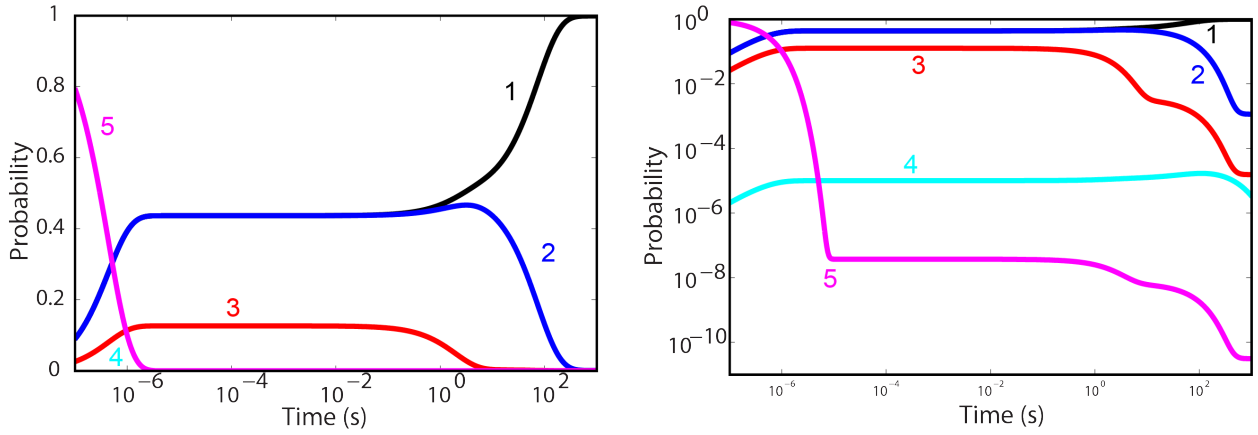

Fig. S5 Predicated time-dependent probability of the different folding states of pIL at zero force. The probability is calculated using the kinetic model and parameters shown in **Fig. 3**. The protein is fully unfolded at the time zero, or $p_5 = 1$. To show the multiple timescales of the folding process, the probability is plotted in both linear scale (left) and logarithm scale (right). The MATLAB code used for this calculation is attached below.

```
% This program calculates the evolution of the different folding states of
% pIL using parameters derived from single-molecule experiments. Code
% written by Dr. Yongli Zhang, Oct., 2011.

q=zeros(5);  % Reaction rates
q(1,5)=3.1e-5; % Transition rate from state 1 to state 2, or unfolding rate
q(2,5)=0.027;
q(3,5)=0.58;
q(4,5)=2.1e-3;
q(5,1)=1e6; q(5,2)=1e6; q(5,3)=2.9e5; q(5,4)=23; % folding rates

ld=diag(true(5,1));  % True for diagonal element
q(ld)=-sum(q,2);

t0=1e-7;
t1=1000;
N=10000;
t=linspace(log(t0),log(t1),N);
t=exp(t);
p=zeros(N,5); % State probability
for i=1:N
   transmat=expm(q.*t(i)); % Solve the master equation.
   p(i,:)=transmat(5,:);
end

% Plot the probability-time data
figure
loglog(t,p,'-')
xlabel('Time (s)')
ylabel('Prob')
```

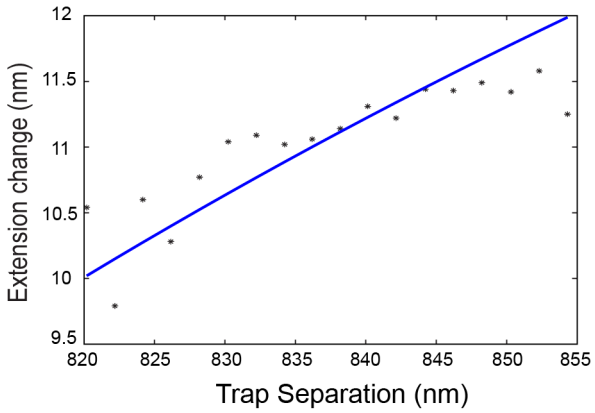

Fig. S6 Trap separation-dependent extension change between the unfolded and folded states of pER (symbols) and their best-fit (line).

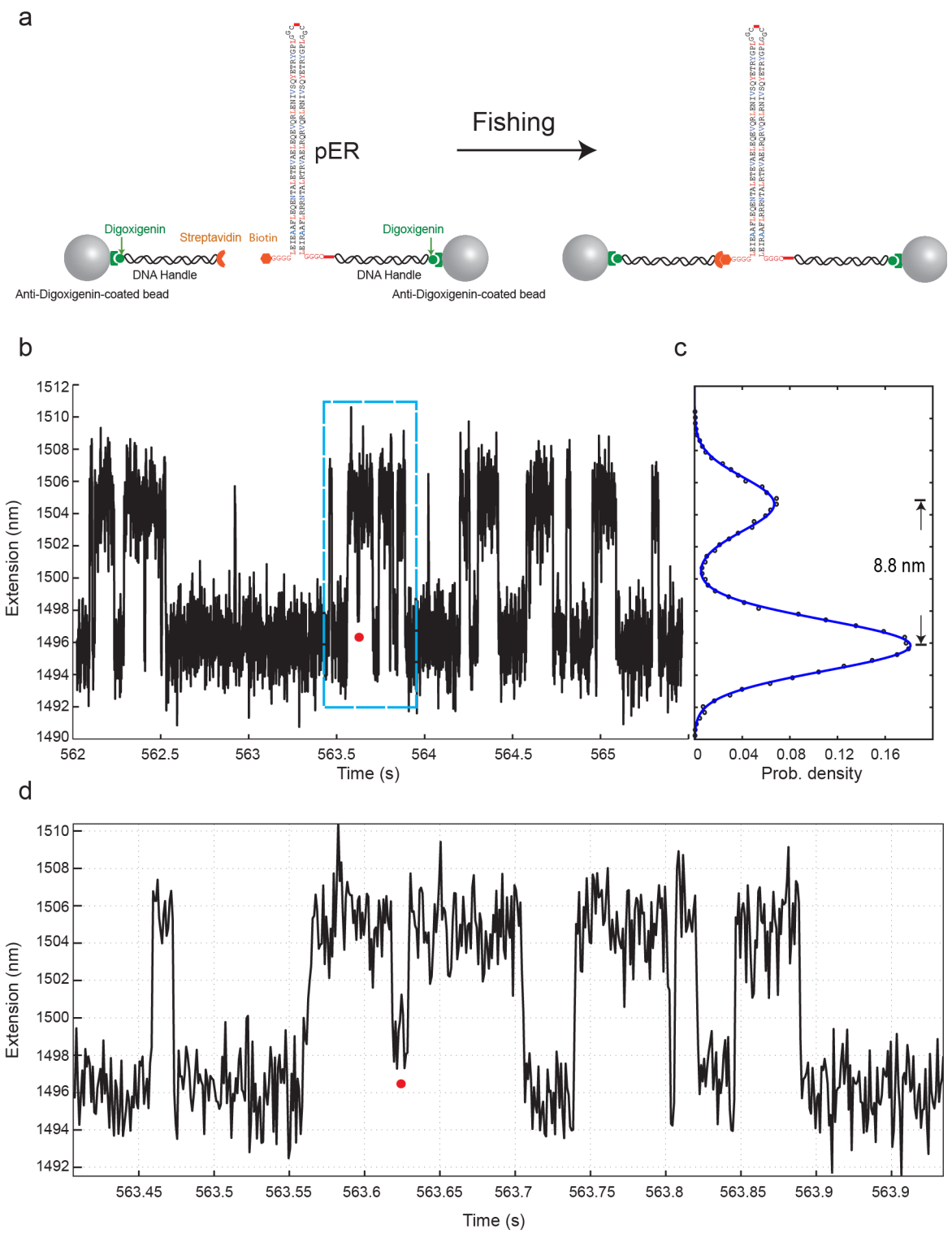

Fig. S7 A single-protein (pER) was pulled using two DNA handles to minimize potential nonspecific protein interaction with bead surfaces. (**a**) The experimental setup. The crosslinked protein-DNA conjugate was first attached to one bead and then brought close to another tapped bead. This second bead was attached with a DNA molecule bound with a streptavidin

molecule. The biotinylated protein was fished and captured by the streptavidin to establish a protein-DNA tether between the two beads, which can be pulled to higher forces. (**b**) The extension-time trace mean-filtered to 1 kHz. A misfolded state is marked by a red dot. (**c**) The probability density distribution of the extension corresponding to that shown (**b**). (**d**) The close-up view of part of trace shown in **c**. The extension traces obtained using two-DNA handles show minor differences compared with those in the one-handle cases, including the decreased average extension change and reduced signal-to-noise ratio. These differences are caused by the increased DNA compliance due to the additional DNA handle(13-15), which can be accounted for by our model (6).

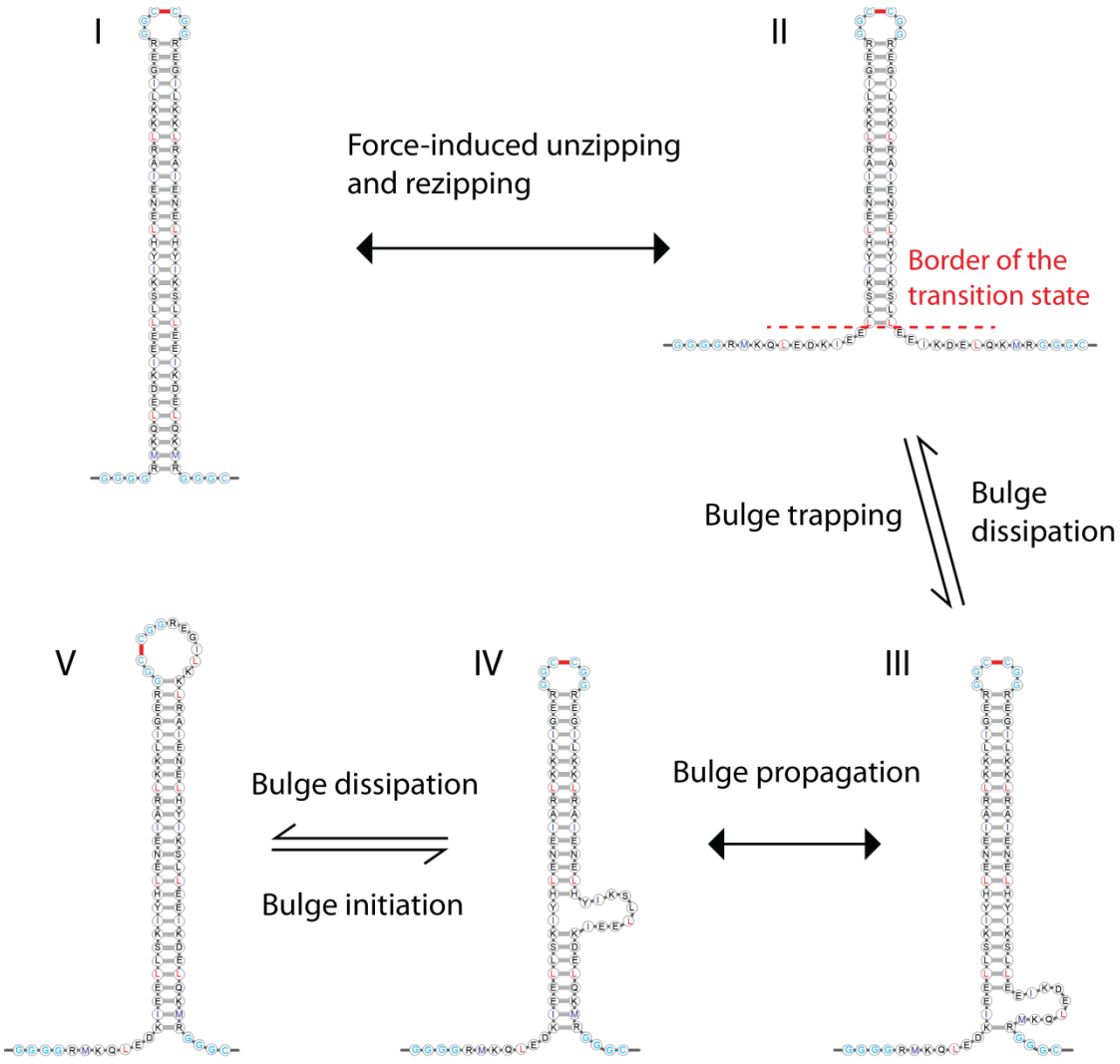

Fig. S8 Possible reptation mechanism of helix sliding in coiled coil proteins. The coiled coil (State I) first unzips from one end due to thermal fluctuations, which can be enhanced by the pulling force (State II). According to the transition state theory, most of these fluctuations lead to only small frayed ends and cannot reach the transition state for productive protein

unfolding. Yet after these small unzipping fluctuations, the unzipped polypeptides may zip with shifted helical registry, trapping one or more heptads of the polypeptide bulge in the alternatively zipped coiled coil (State III). The bulge may diffuse along the helical axis of the coiled coil (through State IV) and dissipate to the other end of the coiled coil, leading to a staggered coiled coil (State V). This reptation mechanism may complete with the misfolding mechanism to form the staggered coiled coil. All the steps illustrated here may be reversible, leading to complete coiled coil folding from the staggered state. Similar reptation mechanisms have been proposed to account for sliding of double-stranded DNA along surfaces of the histone octamers (16, 17) and single-stranded DNA (ssDNA) on surfaces of the ssDNA-binding proteins (18).

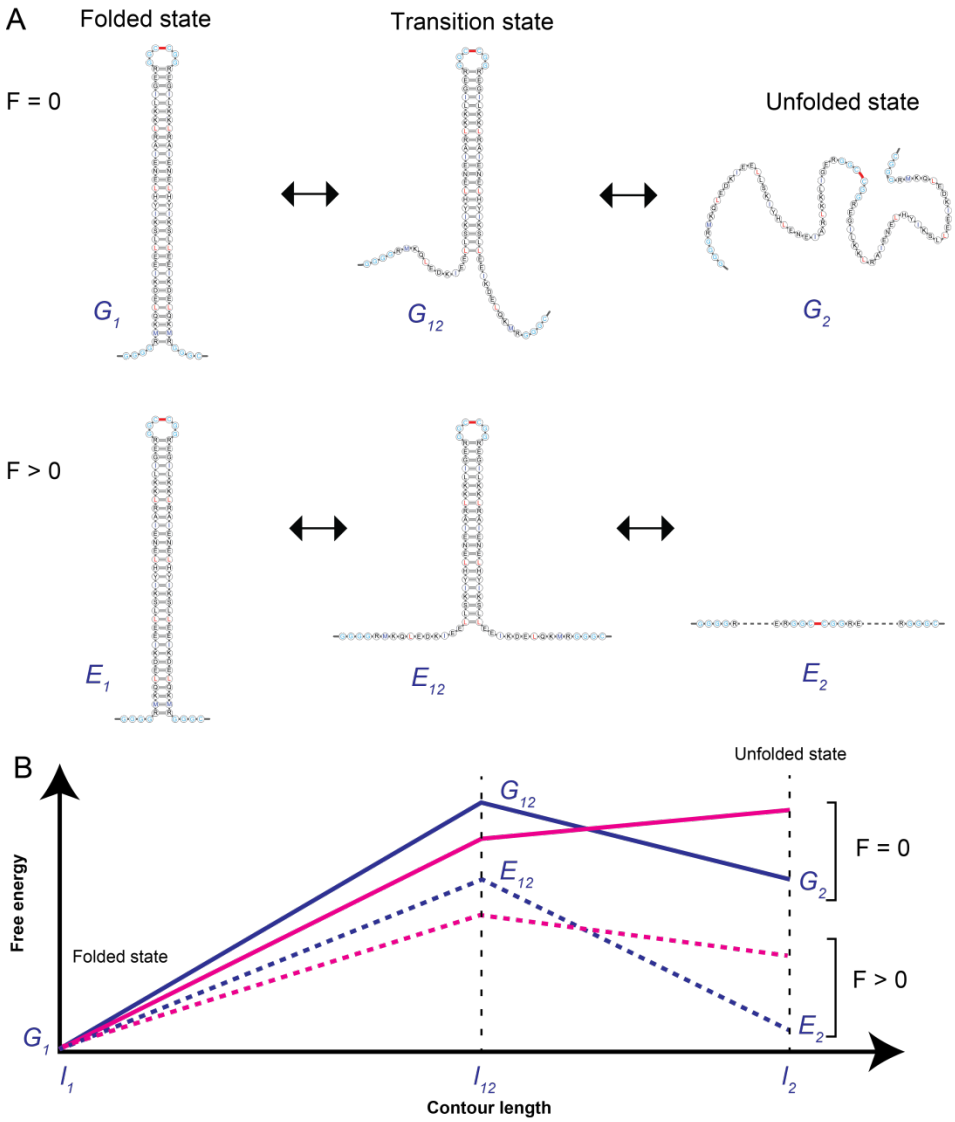

Fig. S9 Effects of force on protein folding kinetics. (**A**) Three characteristic states involved in protein folding in the absence (top row) or presence (bottom row) of force. The folding progress can be described by the contour length of the unfolded polypeptide portion of the protein. (**B**) Force stabilizes both the transition state and the unfolded state relative to the folded state. The solid lines represent the two hypothetical energy landscapes in the absence of force, and the dashed lines are their corresponding landscapes in the presence of force. Note that a down-hill protein folding (indicated by the red solid line) may be required to account for the observed barrier-limited protein folding in the presence of force (red dashed line). For the down-hill folding, the transition state become virtual and is the transition state in the presence of force. In our theory to derive the characteristic energy and structures described in the **SI Text**, we have assumed that positions of these structures in the reaction coordinate are not

shifted by force, an assumption widely used in previous analyses of the data from single-molecule manipulation experiments (6, 19, 20).

References

1. Ainavarapu RK*, et al.* (2007) Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophys J* 92:225-233.
2. Wuite GJ, Smith SB, Young M, Keller D, Bustamante C (2000) Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature* 404:103-106.
3. Comstock MJ, Ha T, Chemla YR (2011) Ultrahigh-resolution optical trap with single-fluorophore sensitivity. *Nat Methods* 8:335-340.
4. Marko JF, Siggia ED (1995) Stretching DNA. *Macromolecules* 28:8759-8770.
5. Bustamante C, Marko JF, Siggia ED, Smith S (1994) Entropic Elasticity of Lambda-Phage DNA. *Science* 265:1599-1600.
6. Gao Y, Sirinakis G, Zhang YL (2011) Highly anisotropic stability and folding kinetics of a single coiled coil protein under mechanical tension. *J Am Chem Soc* 133:12749–12757.
7. Yang WY, Gruebele M (2003) Folding at the speed limit. *Nature* 423:193-197.
8. Husson J, Dogterom M, Pincet F (2009) Force spectroscopy of a single artificial biomolecule bond: The Kramers' high-barrier limit holds close to the critical force. *J Chem Phys* 130.
9. Gebhardt JCM, Bornschlogla T, Rief M (2010) Full distance-resolved folding energy landscape of one single protein molecule. *Proc Natl Acad Sci USA* 107:2013-2018.
10. Cho SS, Weinkam P, Wolynes PG (2008) Origins of barriers and barrierless folding in BBL. *Proc Natl Acad Sci USA* 105:118-123.
11. Meisner WK, Sosnick TR (2004) Barrier-limited, microsecond folding of a stable protein measured with hydrogen exchange: Implications for downhill folding. *Proc Natl Acad Sci USA* 101:15639-15644.
12. Oshea EK, Klemm JD, Kim PS, Alber T (1991) X-ray structure of the GCN4 leucine zipper, a 2-stranded, parallel coiled coil. *Science* 254:539-544.
13. Wen JD, Manosas M, Li PTX, Smith SB, Bustamante C, Ritort F, Tinoco I (2007) Force unfolding kinetics of RNA using optical tweezers. I. Effects of experimental variables on measured results. *Biophys J* 92:2996-3009.
14. Woodside MT, Anthony PC, Behnke-Parks WM, Larizadeh K, Herschlag D, Block SM (2006) Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* 314:1001-1004.
15. Moffitt JR, Chemla YR, Izhaky D, Bustamante C (2006) Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proc Natl Acad Sci USA* 103:9006-9011.
16. Zhang YL, Smith CL, Saha A, Grill SW, Mihardja S, Smith SB, Cairns BR, Peterson CL, Bustamantel C (2006) DNA translocation and loop formation mechanism of chromatin remodeling by SWI/SNF and RSC. *Mol Cell* 24:559-568.

17. Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem* 78:273-304.
18. Zhou R, Kozlov AG, Roy R, Zhang J, Korolev S, Lohman TM, Ha T (2011) SSB functions as a sliding platform that migrates on DNA via reptation. *Cell* 146:222-232.
19. Bustamante C, Chemla YR, Forde NR, Izhaky D (2004) Mechanical processes in biochemistry. *Annu Rev Biochem* 73:705-748.
20. Woodside MT, Behnke-Parks WM, Larizadeh K, Travers K, Herschlag D, Block SM (2006) Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc Natl Acad Sci USA* 103:6190-6195.