Supporting Information

Stigler and Rief 10.1073/pnas.1201801109

SI Text

SI Methods. ***Instrument.*** Experiments were performed on a custom-built passive dual beam optical tweezers setup with one AOD-steerable beam and back focal plane detection based on a setup described earlier (1). Data were collected at 20 kHz bandwidth. Trap calibration was performed using a method introduced by Tolić-Nørrelykke et. al. (2) with all relevant corrections to the power spectrum (3). The error of trap stiffness calibration was approximately 10%.

Construct design. Following a protocol based on a method by Cecconi et al. (4), protein molecules were expressed in *E. coli* with N and C terminal ubiquitin spacers with terminal cysteines that were fused to single stranded DNA oligos using thiol bonds. These protein-DNA hybrids were annealed to dsDNA handles with single stranded overhangs complementary to the hybrid oligos. The dsDNA handles had a total length of about 360 nm and were bound with biotin/neutravidin and digoxigenin/anti-digoxigenin interactions to functionalized silica beads with a diameter of 1 μm. A full description of construct assembly can be found in the Methods section of (5). The full amino acid sequence for the N-terminal domain (construct CaM_{12}) was MACKMQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGELADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTIDFPEFLTMMARKMKGTMQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGKCLE-His_6.

The full sequence for the C-terminal domain (construct CaM_{34}) was MACKMQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGELKMKDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMIREADIDGDGQVNYEEFVQMMTAKGTMQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGKCL-His_6.

Control experiments without ubiquitin spacers yielded comparable experimental results.

Measurements were performed in 50 mM Tris, 150 mM KCl, 26 U/ml glucose oxidase, 17000 U/ml catalase, 0.65% glucose, pH 8.0 with varying concentrations of $CaCl_2$ or EDTA.

We measured multiple molecules at each calcium concentration. The data reported in this study are weighted averages over all molecules at a respective concentration. The number of independent measurements is shown in Table S1.

Data analysis. Data traces were analyzed using a Hidden-Markov classifier (6) applied to the 20 kHz raw data. In brief, the data were coarse grained into about 100 bins. Initial level positions were picked from Gaussian fits to a trace histogram. Subsequent iterations of the forward-backward algorithm and the Baum-Welch algorithm, optimizing the emission probabilities, were applied until only a negligible number of data points was reassigned in an iteration. The quality of the classification was assessed by the distribution of data points in each class and single-exponentiality of lifetimes (See Fig. S1*A*). For a detailed description, see (5, 7).

The Hidden-Markov classifier was not able to robustly separate the apo from the holo state for the C-terminal domain at intermediate calcium concentrations. The resulting lifetime distributions when a two-state model was applied were single exponential for the lower (unfolded) state and double-exponential for the higher (folded) state (see Fig. S1*B*). Since the dwell times for the folded apo and holo conformations are well separated, we determined a threshold from scatter-plots as shown in Fig. 4*C* and assigned all dwells shorter than this threshold to the apo-class (dark green in Fig. 4*C*) and all dwells longer than the threshold to the holoclass (light green in Fig. 4*C*, also see Fig. S1*C*). The dwells in each class followed a truncated cumulative exponential distribution with the CDF:

$$p(t) = \frac{\exp(-k^{\text{off}}t) - \exp(-k^{\text{off}}\tau_{\text{min}})}{\exp(-k^{\text{off}}t) - \exp(-k^{\text{off}}\tau_{\text{min}})}$$

This distribution CDF takes into account that events shorter than a dead time τ_{min} or longer than τ_{max} could not be observed.

Since at low forces, some kinetics approach the temporal resolution of the setup (on the order of a few hundred microseconds), a correction for missed transitions (7) was necessary. The temporal resolution was estimated from the histogram of detected lifetimes.

Linker models. The elasticity of the linker at low forces, where the protein is still folded, was modeled using an extensible worm like chain (eWLC) model (8):

$$F_{\text{eWLC}}(d_{\text{D}}) = \frac{k_{\text{B}}T}{p_{\text{D}}}\left(\frac{1}{4(1-\frac{d_{\text{D}}}{L_{\text{D}}}+\frac{F}{K})^2} - \frac{1}{4} + \frac{d_{\text{D}}}{L_{\text{D}}} - \frac{F}{K}\right).$$

Additional compliances from the unfolded polypeptide chain after unfolding were modeled with a worm like chain model (9) in series with the linker eWLC:

$$F_{\text{WLC}}(d_{\text{P}}) = \frac{k_{\text{B}}T}{p_{\text{P}}}\left(\frac{1}{4(1-\frac{d_{\text{P}}}{L_{\text{P}}})^2} - \frac{1}{4} + \frac{d_{\text{P}}}{L_{\text{P}}}\right).$$

Force-dependent state probabilities and equilibrium energies. After Hidden-Markov classification, the state probabilities $P_i(F)$ were calculated as the time spent in a particular state i divided by the trace time.

The free energy of the complete system is given by

$$\begin{aligned} G_i(F_i) &= G_i^0 + G_i^{\text{device}}(F_i) \\ &= G_i^0 + G^{\text{bead}}(F_i) + G^{\text{linker}}(F_i) + G_i^{\text{p}}(F_i), \end{aligned}$$

where G_i^0 is the free energy of the protein in state i, $G^{\text{bead}}(F) = 1/2 \times (F) \cdot F$ is the energy stored in the displacement x of the beads from the trap center, G^{linker} is the energy stored in the stretching of the linker, and G_i^{p} is the free energy of the unfolded polypeptide. G^{linker} and G_i^{p} are readily calculated as integrals over WLC and eWLC functions. We can now extract the equilibrium free energy of folding ΔG_{ij}^0 between states i and j from a global fit to the force-dependent state probabilities using

$$P_i(F) = \frac{1}{1 + \sum_{j \neq i} \exp(-\frac{\Delta G_{ij}^0 + \Delta G_{ij}^{\text{device}}(F,F_j)}{k_{\text{B}}T})}.$$

For details, see methods in (5). The experimental uncertainty of this procedure is dominated by the uncertainty in trap stiffness determination (about 10%).

An example for a force-dependent probability plot is shown in Fig. S3.

Extraction and extrapolation of rates. For determining rates at zero force and distances to the transition state, the unfolding rates were fitted with a Bell model:

$$k_{\mathrm{unf}}(F) = k_{0,\mathrm{unf}} \exp\left(\frac{F \cdot \Delta x_{\mathrm{unf}}}{k_{\mathrm{B}} T}\right).$$

Folding rates were fitted with a model accounting for the energy differences in the DNA linker and beads deflection between unfolded state and transition state T (10):

$$k_{\mathrm{fold}}(F = F_i) = k_{0,\mathrm{fold}} \exp\left(-\frac{\Delta G_{i\mathrm{T}}(F_i, F_{\mathrm{T}})}{k_{\mathrm{B}} T}\right).$$

Model for folding and unfolding. In an equilibrium model, the folding/unfolding can be described in a scheme as shown in Fig. 6 (11, 12). In this model it is assumed that folding occurs via a transition state that can bind up to two calcium ions in equilibrium. The barrier crossing time (i.e. folding transitions) are assumed to be faster than the equilibrium of the transition state.

The probability to find the system in a state with no calcium bound is given by

$$P(N_0) = \frac{[N_0]}{[N_0] + [N_1] + [N_2]} = \frac{1}{1 + \frac{[L]}{K_{\mathrm{N1}}} + \frac{[L]^2}{K_{\mathrm{N1}} K_{\mathrm{N2}}}},$$

where K_{N1} and K_{N2} are the macroscopic dissociation constants for the first and second calcium ion, respectively, and $[L]$ is the calcium concentration.

Likewise, the probability to find the system with one calcium bound or two calcium ions bound is given by $P(N_1) = \frac{1}{1+\frac{K_{\mathrm{N1}}}{[L]}+\frac{[L]}{K_{\mathrm{N2}}}}$ and $P(N_2) = \frac{1}{1+\frac{K_{\mathrm{N2}}}{[L]}+\frac{K_{\mathrm{N1}} K_{\mathrm{N2}}}{[L]^2}}$, respectively.

The effective rate for unfolding is now given by

$$k_u = P(N_0) k_{u,0} + P(N_1) k_{u,1} + P(N_2) k_{u,2},$$

where the intrinsic unfolding rates with one calcium bound is given by

$$k_{u,1} = k_{u,0} \exp\left(-\frac{\Delta G_{\mathrm{N1}} - \Delta G_{\mathrm{TS1}}}{k_{\mathrm{B}} T}\right) = k_{u,0} \frac{K_{\mathrm{N1}}}{K_{\mathrm{TS1}}},$$

and the intrinsic unfolding rate with two calcium ions bound is given by

$$k_{u,2} = k_{u,1} \exp\left(-\frac{\Delta G_{\mathrm{N2}} - \Delta G_{\mathrm{TS2}}}{k_{\mathrm{B}} T}\right) = k_{u,1} \frac{K_{\mathrm{N1}}}{K_{\mathrm{TS1}}}$$
$$= k_{u,0} \frac{K_{\mathrm{N1}} K_{\mathrm{N2}}}{K_{\mathrm{TS1}} K_{\mathrm{TS2}}}.$$

Similarly, the effective folding rate is given by

$$k_{\mathrm{f}} = P(N_0) k_{\mathrm{f},0} + P(N_1) k_{\mathrm{f},1} + P(N_2) k_{\mathrm{f},2},$$

with

$$k_{\mathrm{f},1} = k_{\mathrm{f},0} \frac{K_{\mathrm{U1}}}{K_{\mathrm{TS1}}} \quad \text{and} \quad k_{\mathrm{f},2} = K_{\mathrm{f},0} \frac{K_{\mathrm{U1}} K_{\mathrm{U2}}}{K_{\mathrm{TS1}} K_{\mathrm{TS2}}}.$$

Taken together, we obtain

$$k_{\mathrm{f}} = k_{\mathrm{f},0} \frac{1 + \frac{[L]}{K_{\mathrm{TS1}}} + \frac{[L]^2}{K_{\mathrm{TS1}} K_{\mathrm{TS2}}}}{1 + \frac{[L]}{K_{\mathrm{U1}}} + \frac{[L]^2}{K_{\mathrm{U1}} K_{\mathrm{U2}}}} \quad \text{and} \quad k_{\mathrm{u}} = k_{\mathrm{u},0} \frac{1 + \frac{[L]}{K_{\mathrm{TS1}}} + \frac{[L]^2}{K_{\mathrm{TS1}} K_{\mathrm{TS2}}}}{1 + \frac{[L]}{K_{\mathrm{N1}}} + \frac{[L]^2}{K_{\mathrm{N1}} K_{\mathrm{N2}}}}.$$

We used $\log_{10} K_{\mathrm{N1}} = -4.62$ and $\log_{10} K_{\mathrm{N2}} = -5.17$ (in molar units) as values for the molar dissociation constants to the native state of CaM_{12} and $\log_{10} K_{\mathrm{N1}} = -5.32$ and $\log_{10} K_{\mathrm{N2}} = -6.21$ for CaM_{34} (13). Additionally, we assumed that the unfolded state is unsaturated at our conditions: $\log_{10} K_{\mathrm{U1}} = log_{10} K_{\mathrm{U2}} = 1$. We performed a global fit to both the folding and unfolding rates at zero force (see Fig. 5 *A–D*). The dissociation constants of the transition state K_{TS1} and K_{TS2} as well as the apo folding rate at zero calcium $k_{f,0}$ and the apo unfolding rate at zero calcium $k_{u,0}$ were free fit parameters.

We find for CaM_{12} $\log_{10} K_{\mathrm{TS1}} = -3.6 \pm 0.2$, $\log_{10} K_{\mathrm{TS2}} > 1$ in molar units, and $\log_{10} k_{\mathrm{f},0} = 4.2 \pm 0.1$ and $\log_{10} k_{\mathrm{u},0} = 0.6 \pm 0.1$ in units of s^{-1}. For CaM_{34} we find $\log_{10} K_{\mathrm{TS1}} = -5.4 \pm 0.1$, $\log_{10} K_{\mathrm{TS2}} > 1$, $\log_{10} k_{\mathrm{f},0} = 2.8 \pm 0.1$ and $\log_{10} k_{\mathrm{u},0} = 1.9 \pm 0.1$.

Equilibrium free energy under ligand binding conditions. The equilibrium free energies of folding can be calculated according to

$$-\frac{\Delta G_0}{k_{\mathrm{B}} T} = -\frac{\Delta G_0^{[L]=0}}{k_{\mathrm{B}} T} - \ln \frac{1 + \frac{[L]}{K_{\mathrm{N1}}} + \frac{[L]^2}{K_{\mathrm{N1}} K_{\mathrm{N2}}}}{1 + \frac{[L]}{K_{\mathrm{U1}}} + \frac{[L]^2}{K_{\mathrm{U1}} K_{\mathrm{U2}}}}$$
$$\cong -\frac{\Delta G_0^{[L]=0}}{k_{\mathrm{B}} T} - \ln \frac{[L]^2}{K_{\mathrm{N1}} K_{\mathrm{N2}}}.$$

The continuous lines in Fig. 5 *E* and *F* are fits to this function. The fit parameters of $-\Delta G_0^{[L]=0} = 7.7\, k_{\mathrm{B}} T$ and $log_{10}(K_{\mathrm{N1}} \cdot K_{\mathrm{N2}}) = -9.2$ for CaM_{12} correspond well to values measured by chemical denaturation of $4.5\, k_{\mathrm{B}} T$–$6.4\, k_{\mathrm{B}} T$ (14) and –9.8 (13). For CaM_{34} the fit parameters are $-\Delta G_0^{[L]=0} = 2.5\, k_{\mathrm{B}} T$ and $log_{10}(K_{\mathrm{N1}} \cdot K_{\mathrm{N2}}) = -11.5$, compared with literature values of $3.5\, k_{\mathrm{B}} T$–$4.5\, k_{\mathrm{B}} T$ (14) and –11.5 (13).

1. Gebhardt JCM, Bornschlögl T, Rief M (2010) Full distance-resolved folding energy landscape of one single protein molecule. *PNAS* 107:2013–2018.
2. Tolić-Nørrelykke S et al. (2006) Calibration of optical tweezers with positional detection in the back focal plane. *Rev Sci Instrum* 77:103101–103112.
3. Berg-Sørensen K, Flyvbjerg H (2004) Power spectrum analysis for optical tweezers. *Rev Sci Instrum* 75:594–612.
4. Cecconi C, Shank EA, Dahlquist FW, Marqusee S, Bustamante C (2008) Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. *Eur Biophys J* 37:729–738.
5. Stigler J, Ziegler F, Gieseke A, Gebhardt JCM, Rief M (2011) The complex folding network of single calmodulin molecules. *Science* 334:512–516.
6. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE* 77:257–286.
7. Stigler J, Rief M (2012) Hidden markov analysis of trajectories in single-molecule experiments and the effects of missed events. *Chemphyschem* 13:1079–1086.
8. Wang MD, Yin H, Landick R, Gelles J, Block SM (1997) Stretching DNA with optical tweezers. *Biophys J* 72:1335–1346.
9. Bustamante C, Marko JF, Siggia ED, Smith SB (1994) Entropic elasticity of lambda-phage DNA. *Science* 265:1599–1600.
10. Schlierf M, Berkemeier F, Rief M (2007) Direct observation of active protein folding using lock-in force spectroscopy. *Biophys J* 93:3989–3998.
11. Bodenreider C, Kiefhaber T (2005) Interpretation of protein folding psi values. *J Mol Biol* 351:393–401.
12. Fersht AR (2004) Phi value versus psi analysis. *PNAS* 101:17327–17328.

13. Bayley PM, Findlay WA, Martin SR (1996) Target recognition by calmodulin: Dissecting the kinetics and affinity of interaction using short peptide sequences. *Protein Sci* 5:1215–1228.

14. Masino L, Martin SR, Bayley PM (2000) Ligand binding and thermodynamic stability of a multidomain protein, calmodulin. *Protein Sci* 9:1519–1529.

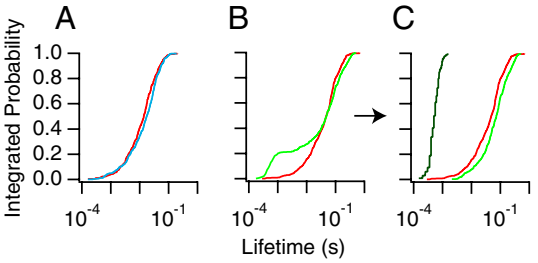

Fig. S1. Integrated lifetime histograms of the 100 μM Ca^{2+} traces shown in the middle panels of Figs. 3*B* and 4*B*. (*A*) Integrated histogram for CaM_{12}. (*B*) Integrated histogram for CaM_{34}. (*C*) Integrated histogram of (*B*) after separating the short-lived state (now shown in dark green) from the long-lived one. The data in (*A*) and (*C*) follow single exponential distributions.

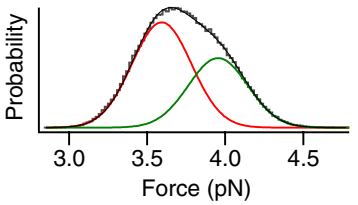

Fig. S2. Distribution of force values of the trace shown in Fig. 4*B*, *Right* (grey). The histogram can be fitted (black) with two gaussians for the unfolded state (red) and the folded state (green).

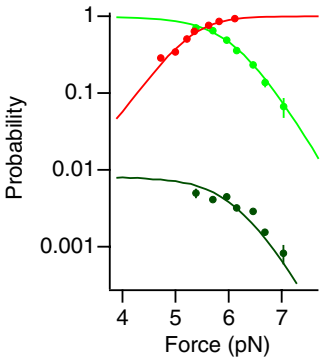

Fig. S3. Force-dependent probabilities for the isolated C-terminal domain (CaM_{34}) at 10 μM Ca^{2+} conditions. Shown are the force-dependent probabilities to find the molecule in the unfolded state (red), the folded holo state (light green), and the folded apo state (dark green) together with a global fit to obtain the equilibrium free energies.

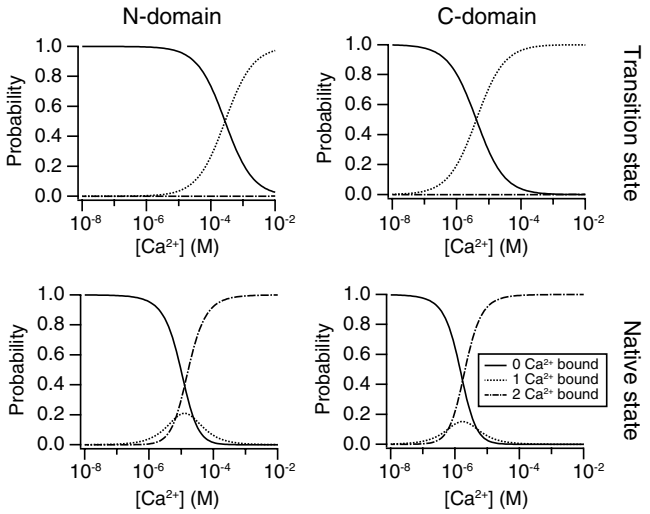

Fig. S4. Equilibrium populations of the states in the scheme shown in Fig. 6 as calculated from the dissociation constants (see *SI Text*) at varying calcium concentrations for the N-terminal domain (CaM_{12}) and the C-terminal domain (CaM_{34}).

Table S1. Number of molecules measured at each calcium concentration for the N-terminal domain (CaM_{12}) and the C-terminal domain (CaM_{34})

Calcium concentration (mM)	Number of molecules	
	CaM_{12}	CaM_{34}
0 [10 mM EDTA]	2	4
~0.01	1	7
0.06	2	5
0.1	6	5
0.8	3	1
10	3	4