SI Text

Synthesis, Expression, and Purification of CpLC. A synthetic gene encoding 63 aa of the Ig-binding domain of protein L and flanked by Cys residues (termed CpLC, $C^{1}EEV^{4}\ldots.F^{62}AGC^{65}$) was cloned between the *Nde*I and BamH1 restriction endonuclease sites of pET11a vector (Novagen, San Diego, CA) and expressed in *E.coli* BL21(DE3; Stratgene, La Jolla, CA). Cells were grown at 37°C in Luria-Bertani medium, and expression was induced at 0.7 OD_{600} with a final concentration of 1 mM isopropylthiogalactoside for a period of 3-4 h. Typically, cells harvested from a 400-ml culture were suspended in 16-20 ml of 1X PBS (1.7 mM KH_2PO_4, 5 mM Na_2HPO_4, 150 mM NaCl, pH 7.4), heated at 80°C for 5 min, immediately chilled on ice for 10 min, followed by centrifugation at 16,000 rpm (SS-34 rotor, Sorvall) for 30 min at 4°C. The supernatant was passed although a 0.45 μM syringe filter, dialyzed against an excess of 0.5X PBS and 2 mM DTT (DTT) in the cold room, concentrated using centriprep YM-3 devices (Millipore Corp, Bedford, MA), and loaded onto a Superdex-75 column (2.6 cm × 60 cm; GE HealthCare, Piscataway, NJ) equilibrated in the same buffer at a flow rate of 3 ml/min at room temperature. Peak fractions were combined and subjected to reverse-phase HPLC chromatography on POROS 20 R2 resin (Perceptive Biosystems, Framingham, MA) and eluted using a linear gradient of 0% to 60% acetonitrile/0.05% trifluoroacetic acid/water over a period of 16 min at a flow rate of 4 ml/min. Aliquots (0.5 mg) of the peak fraction were lyophilized and stored at –70°C. The CpLC construct was verified both by DNA sequencing and mass spectrometry.

Labeling of CpLC with Alexa Dyes. CpLC (0.5 mg) was dissolved in 500 μL of degassed 50 mM sodium phosphate buffer (pH 7) to achieve a final concentration of ~140 μM in an oxygen-free chamber. One hundred micrograms of Alexa 488 maleimide and Alexa 594 maleimide dyes (Molecular Probes, Carlsbad, CA) each dissolved in 20 μl of DMSO were added to the protein solution and stirred for 1 h in the dark at room temperature under nitrogen atmosphere. Subsequently, the protein solution was incubated with 20 mM 2-mercapto-ethanol for 1-2 h at room temperature to prevent any unreacted dye from binding to the resin in the subsequent column steps. The reaction mixture was

first loaded onto a Superdex-30 column (1.6 cm x 60 cm) equilibrated in 0.5× PBS and 1 mM DTT to separate the unreacted dye from the labeled protein. Peak fractions were then fractionated on a Mono-Q column (1 ml, GE HealthCare) using a 0-1 M NaCl gradient in 20 mM Tris-HCl, pH 8 and 1 mM DTT, over a period of 30 min at a flow rate of 1 ml/min at room temperature monitored at threespecific wavelengths, two specific to the dye and one at 280 nm. Peak fraction corresponding to the protein having an equimolar ratio of the two Alexa dyes was concentrated and run again on the same Superdex column in 20 mM $NH_4C_2H_3O_2$ (pH 6.7) to remove the salt and DTT.

Sample Preparation and Measurement. Protein samples were diluted to 35 pM in GdmCl solutions containing 0.01% Tween 20 and buffered with 50 mM sodium phosphate (final pH 7.0). All GdmCl concentrations were determined from the refractive index.

Single molecule FRET efficiencies for protein L and Csp*Tm* were measured on a Picoquant (Berlin, Germany) Microtime 200 fluorescence microscope with time-correlated single-photon counting capabilities using a 100 × 1.4 N.A. Olympus PlanApo objective. The donor dye was excited using a 470-nm pulsed laser (Picoquant, LDH-470, 20 MHz repetition rate, 80 ps FWHM, 35 μW average power). After being focused through a 75 μm pinhole, the donor and acceptor fluorescence was separated with a dichroic beam splitter (Chroma 585 DCXR) and filtered (donor channel: Omega 525AF45; acceptor channel: Omega 600 ALP) before being detected by single-photon counting avalanche photodiodes (Perkin-Elmer Optoelectronics SPQM-AQR15). The arrival time of each photon (100-ns resolution) as well as the fluorescence delay time relative to laser pulse (37-ps resolution) were recorded for each detection channel and stored for later analysis. Sample drops were sealed between a quartz cover glass and an imaging chamber cover (Grace Biolabs). Data were collected from a sample for 8-24 h, and multiple samples were combined to obtain up to 70 h of data at each denaturant concentration.

Fitting the FRET Histograms. The measured FRET efficiency for each burst was computed as $E_m = (n_A - n_D l)/(n_A + n_D)$, with donor leakage probability l varying between $0.053 \le l = 0.067$ depending on denaturant concentration. The folded and unfolded populations of protein L are not clearly resolved because of the comparatively large distance between the carboxyl and amino termini of the folded state. There is also a peak near $E \sim 0$ associated with proteins lacking a photoactive acceptor. The FRET histograms were well fit with three Gaussian functions (for the folded, unfolded, and $E \sim 0$ populations) for GdmCl concentrations between 1.5 M and 3.25 M, and with two Gaussian functions at all other denaturant concentrations. The FRET efficiency of the folded protein changes even though the average distance between the two dyes in the folded state is independent of the GdmCl concentration because both γ and R_0 vary with denaturant concentration. To improve the reliability of the fitting for 1.5 M – 3.25 M GdmCl, the donor-acceptor distance measured for the folded state in 0 M GdmCl was used to compute $\langle E_m \rangle$ for folded states at all denaturant concentrations using the appropriate values of γ and R_0. The Förster radius R_0 was determined at each denaturant concentration by measuring the overlap between donor fluorescence and acceptor absorption spectra, donor quantum yield, and the refractive index of the bulk solvent (1). $\kappa^2 = 2/3$ was used throughout the analysis, as justified by the Langevin simulations.

The fluorescence anisotropy for Csp*Tm* has been reported previously by Schuler *et al.* to be between 0.03 and 0.09 (2). The steady state fluorescence anisotropy, *A*, for folded protein L measured with a spectrofluorometer (Spex Fluorolog 2) was found to be 0.05 ± 0.02, in agreement with a similar dye-labeled protein L construct used by Sherman and Haran (3), and consistent with the value of 0.09 estimated from

$$A \equiv \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} = \frac{3\cos^2\theta - 1}{5\left(1 + \tau_D/\tau_c\right)} \qquad [1]$$

using the measured donor life-time (τ_D) of 1.0 ns, the reorientational correlation time (τ_c) of 0.3 ns predicted from the Langevin simulations, and assuming $\theta = 0$ for the angle between the absorption and emission transition dipole moments.

The folded and unfolded populations of Csp*Tm* are easily resolved in the FRET histograms. The FRET efficiency histograms were well fit by two Gaussian functions for GdmCl concentrations of 1–3 M, where both folded and unfolded populations are present, and to a single Gaussian function at all other concentrations. The widths of the folded peaks in the presence of GdmCl were set equal to the width of the folded state at 0 M GdmCl.

Langevin and Molecular Dynamics Simulations. Langevin simulations of a simplified model for the polypeptide structure were used to estimate the folded FRET efficiency, and the relative timescales of dye and chain motion, in which each amino acid residue is represented as a spherical bead centered at the α-carbon position with a mass proportional to its size (Fig. 1). For the protein the Gō-like energy function (i.e., where the only attractive interactions are between residues that form contacts in the native structure) of Karanicolas and Brooks was used (4). The linkers were coarse-grained on a similar length scale, each being represented by four beads connected by bonds of length 3.8 A, and the dyes by three beads bonded together (the coarse linker and "dye" particles have radii 3.0 and 4.5 Å, respectively) (Fig. 1). The torsional potential of the linker was taken from the Gly-Gly pseudodihedral in the protein model (4). The interaction of the dyes with each other and with the protein was taken as purely repulsive. The transition dipole moments lie along an axis described by two of the coarse dye particles, corresponding to the long axes of the dyes. This model will be denoted "Gō-dye" to distinguish it from the other coarse simulations presented below. Simulations of length 0.75 μs were run with Langevin dynamics for both the folded and unfolded protein with friction of 50 ps^{-1} at a temperature of 300 K using the CHARMM code (5); all but the dyes and the last three amino acids at either end were fixed for the folded simulations.

Because both dyes are negatively charged, we examined the influence of including charge for the simulation of folded *CspTm*. Adding the net charges to each residue of the protein and to the dyes (-1 on the Alexa 488 donor and -1 on the Alexa 594 acceptor), with a dielectric constant of 80 and Debye screening length of 10 Å, we obtain very similar results for the end-to-end distribution and FRET efficiency (0.99). Using unscreened charges on the full system or on the dyes alone (which is unphysical because the measurements are made in 50 mM sodium phosphate buffer) reduced the efficiency to 0.95.

All-atom molecular dynamics simulations of the unfolded proteins were run using the OPLS-AA/L force-field and the GROMACS 3.3 simulation package. Initial configurations were generated from simulations in implicit solvent at 800 K in CHARMM. The same set of five initial configurations was solvated in a 60 Å solvent box: the solvent was either water or a mixture of water and urea of concentration ≈2.1, 4.2, or 6.2 M. Ten sodium and seven chloride ions were added to neutralize the total charge, and electrostatics were treated by a generalized reaction field technique. The simulations were run at constant temperature and pressure using a Nose–Hoover thermostat and Parrinello–Rahman barostat for 20-25 ns per initial configuration; the first 5 ns was discarded before analysis.

Calculation of Fluorescence Decay from Simulations. Simulations of the Gō-dye model were used to calculate the decay of donor fluorescence intensity directly. The decay of donor fluorescence intensity of donor-only labeled protein is assumed to be single-exponential with rate constant k_D (in all of the calculations presented, the donor lifetime, $\tau_D = 1/k_D$, was taken to be 4.2 ns). In the presence of an acceptor, the instantaneous rate of energy transfer is given by:

$$k_{ET} = \alpha \frac{(\kappa(t))^2}{(R(t))^6} \quad [2]$$

where the constant α is $(3/2)k_D R_0^6$ (R_0 is the Förster distance, taken to be 54.0 Å for Alexa 488 and Alexa 594), and the orientational factor κ is given by:

$$\kappa(t) = \hat{\mu}_D(t) \cdot \hat{\mu}_A(t) - 3(\hat{r}_{DA}(t) \cdot \hat{\mu}_D(t))(\hat{r}_{DA}(t) \cdot \hat{\mu}_A(t)) \quad [3]$$

with $\hat{\mu}_D$, $\hat{\mu}_A$ and $\hat{r}_{DA}$ being, respectively, unit vectors along the donor and acceptor dipole moments and the vector between the donor and acceptor positions. For comparison, the calculations below were also done using $\kappa^2 = 2/3$, the correct value if complete (isotropic) orientational averaging occurs over times much shorter than the donor fluorescence lifetime.

The decay of fluorescence intensity (fluorescence life-time distribution) was computed from the instantaneous transfer rates using (6):

$$I(t) = \left\langle \exp\left(-\int_0^t [k_D + k_{ET}(\tau)]\, d\tau \right) \right\rangle \quad [4]$$

Averaging was done over different time origins regularly spaced throughout the trajectory of the Gō-dye model for time lags *t* of up to 20 ns. FRET efficiencies were computed directly from the relation:

$$\langle E \rangle = 1 - k_D \int_0^\infty I(t) dt \quad [5]$$

In addition to the direct calculation, the efficiencies were also calculated in two theoretical limits (7) of (*i*) fast orientational dynamics, slow chain dynamics, and (*ii*) fast orientational dynamics, fast chain dynamics.

If the linker dynamics (or unfolded chain and linker dynamics for unfolded proteins) is much slower and complete oriental averaging of the dyes is much faster than the fluorescence life-time, averaging occurs over a static distribution of donor-acceptor distances R given by the radial distribution function $P_{eq}(R)$. In this case, the donor fluorescence decay $I(t)$ is computed as

$$I(t) = \int_0^L P_{eq}(R) \exp\left[-k_D\left(1+\left(\frac{R_0}{R}\right)^6\right)t\right] dR \quad [6]$$

and the mean efficiency is

$$\langle E \rangle = \int_0^L \frac{P_{eq}(R)}{1+(R/R_0)^6} dR \quad . [7]$$

If both orientational averaging and chain reconfiguration occur on times shorter than the fluorescence life-time, the mean efficiency is given by:

$$\langle E \rangle = \frac{\int_0^L (R_0/R)^6 P_{eq}(R) dR}{1 + \int_0^L (R_0/R)^6 P_{eq}(R) dR} \quad . [8]$$

Bead Model for Unfolded Chains. A second coarse model (which we will call the "bead model") was used to simulate the effects of excluded volume and attractive interactions in the denatured state. This bead model was used to fit the mean FRET efficiencies and compute donor fluorescence decays. The chain consists of 72 beads (representing the 64 residues of protein L and linkers; the same description was used for Csp*Tm*) with bond lengths constrained to 3.81 Å using SHAKE and harmonic bond angles [$\theta_0 = 91.7°$; $k_\theta =$ 75.6 kcal/(mol.rad^2)]. Nonbonded terms were treated using an adaptation of the Weeks,

Chandler, and Anderson separation (8) of attractive and repulsive parts of a Lennard–Jones potential.

The CHARMM Lennard–Jones potential is given by

$$U_{LJ}(r) = \varepsilon\left\{\left(\frac{\sigma}{r}\right)^{12} - 2\left(\frac{\sigma}{r}\right)^{6}\right\} \quad [9]$$

which can be separated into an attractive part and repulsive part U_r and attractive part U_a as follows:

$$U_r(r) = \begin{cases} U_{LJ}(r) + \varepsilon & r < \sigma \\ 0 & r \geq \sigma \end{cases} \quad [10]$$

$$U_a(r) = \begin{cases} -\varepsilon & r < \sigma \\ U_{LJ}(r) & r \geq \sigma \end{cases} \quad [11]$$

The attractive and repulsive parts can be combined using

$$U_c(r) = \lambda U_a(r) + U_r(r). \quad [12]$$

When $\lambda = 1$, $U_c = U_{LJ}$, when $\lambda \neq 1$ U_c is still a differentiable potential (up to first derivative). Here, by setting $\varepsilon = 1$ kcal/mol for all interactions, we vary the excluded volume to mimic the increase in chain thickness from bound denaturant by changing σ and the attractive interactions by changing λ. In this way, for a given σ, the repulsive part of the potential is always the same, and the repulsive and attractive components can be varied independently.

Langevin dynamics simulations of length 0.75 μs were run in a modified copy of the CHARMM program (9) at 300 K with a friction of 0.2 ps^{-1}. By varying either σ or λ, the

simulated distributions were matched to experimental mean efficiencies by calculating $\langle E \rangle$ from a $P_{eq}(R)$ distribution from simulation assuming slow chain dynamics, using Eq. **7**.

The appropriate R_0 (described in *Fitting the FRET Histogram*) was used at each denaturant concentration. The fitted parameters for the different models and conditions are listed in Table 1.

Effect of Chain Diffusion Dynamics. Chain diffusion dynamics can affect both the calculated radius of gyration determined by the experimental FRET efficiency and the shape of the donor fluorescence decay. At low laser intensities, the mean FRET efficiency defined using donor and acceptor intensities can be calculated using the steady-state populations of the donor-excited ($\rho(D^*A)$) and acceptor-excited ($\rho(DA^*)$) populations (10)

$$\langle E \rangle = \frac{\langle n_A \rangle}{\langle n_A \rangle + \gamma \langle n_D \rangle} = \frac{k_A \rho(DA^*)}{k_A \rho(DA^*) + k_D \rho(D^*A)} \quad [13]$$

where k_A and k_D are the acceptor and donor relaxation rates. The steady-state populations $\rho(D^*A) = \int p_{D^*A}(R)dR$ and $\rho(DA^*) = \int p_{DA^*}(R)dR$ were found by solving the coupled differential equations for $p_{D^*A}(R)$ and $p_{DA^*}(R)$, which includes population transfer dynamics due to excitation, relaxation, and FRET transfer, as well as diffusion in a one-dimensional potential with reflecting boundary conditions:

$$\begin{aligned} D\frac{\partial}{\partial R}P_{eq}(R)\frac{\partial}{\partial R}\frac{p_{D^*A}}{P_{eq}(R)} - \left(k_D + k_{ET}(R)\right)p_{D^*A} + k_{ex} &= 0 \\ D\frac{\partial}{\partial R}P_{eq}(R)\frac{\partial}{\partial R}\frac{p_{DA^*}}{P_{eq}(R)} + k_{ET}(R)\,p_{D^*A} - k_A p_{DA^*} &= 0 \end{aligned} \quad [14]$$

D is the diffusion coefficient, and the chain end-to-end distribution $P_{eq}(R)$ is assumed to be that for a Gaussian chain random coil. The rate of energy transfer $k_{ET}(R) = k_D (R_0 / R)^6$ assumes fast orientational dynamics. k_{ex} is the excitation rate, which cancels in Eq. **12** because both $\rho(D^* A)$ and $\rho(DA^*)$ are proportional to k_{ex}. The diffusion operator as well as the energy transfer rate were discretized along R. $\rho(D^* A)$ and $\rho(D^* A)$ in Eq. **13** were computed by varying $\langle R^2 \rangle$ of $P_{eq}(R)$ for a given D to match the experimental mean FRET efficiency. Once $P_{eq}(R)$ was determined, the donor fluorescence decay

$$I(t) = \int_0^L p(R,t) dR \qquad [15]$$

was calculated by solving for $p(R,t)$, which satisfies the diffusion equation

$$\frac{\partial}{\partial t} p(R,t) = D \frac{\partial}{\partial R} P_{eq}(R) \frac{\partial}{\partial R} \frac{p(R,t)}{P_{eq}(R)} - \left(k_D + k_{ET}(R)\right) p(R,t) \qquad [16]$$

with initial condition $p(R,0) = P_{eq}(R)$.

Donor fluorescence decays for three different values of D [static, i.e., $D = 0$; the experimentally determined upper bound $D = 16$ Å^2/ns (11); and 10 times the upper bound, $D = 160$ Å^2/ns] are shown in Fig. 8. In general, for a system with a given mean FRET efficiency $\langle E \rangle$, including the effects of chain diffusion dynamics affects the shape of the donor fluorescence decay and increases the computed R_g of the unfolded state. However, as can be seen in Fig. 1, chain diffusion has virtually no effect on the donor fluorescence decay for realistic values of D, as compared with a static chain configuration, and increases R_g less than 3%. It is therefore reasonable to ignore the

effects of diffusion and treat the chain as essentially static on the time scale of the donor life-time.

1. Van Der Meer BW, Coker G, III, Chen SYS (1994) *Resonance Energy Transfer: Theory and Data* (VCH, New York).

2. Schuler B, Lipman EA, Eaton WA (2002) *Nature* 419:743-747.

3. Sherman E, Haran G (2006) *Proc Natl Acad Sci USA* 103:11539-11543.

4. Karanicolas J, Brooks CL (2002) *Protein Sci* 11:2351-2361.

5. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) *J Comp Chem* 4:187-217.

6. Henry ER, Hochstrasser RM (1987) *Proc Natl Acad Sci USA* 84:6142-6146.

7. Schuler B, Lipman EA, Steinbach PJ, Kumke M, Eaton WA (2005) *Proc Natl Acad Sci USA* 102:2754-2759.

8. Weeks JD, Chandler D, Andersen HC (1971) *J Chem Phys* 54:5237-5247.

9. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) *J Comp Chem* 4:187-217.

10. Gopich I, Szabo A (2005) *J Chem Phys* 122:014707.

11. Buscaglia M, Kubelka J, Eaton WA, Hofrichter J (2005) *J Mol Biol* 347:657-664.