Supporting Online Material for

Single molecule fluorescence experiments determine protein folding transition path times

Hoi Sung Chung, Kevin McHale, John M. Louis, and William A. Eaton

correspondence to: chunghoi@niddk.nih.gov or eaton@helix.nih.gov

This PDF file includes:

Materials and Methods
Figs. S1 to S9
Table S1
Full References

Supporting Online Material

Materials and Methods

Expression, purification, and dye-labeling of GB1 and WW domain. The amino acid sequences of the 56-residue long immunoglobulin-binding domain B1 of streptococcal protein G (GB1; PDB code 3GB1) and the formin-binding protein (FBP) WW domains (PDB code 1E0L) are shown in Fig. S1.

GB1 used in this study is flanked by the biotin acceptor peptide (AviTag, Avidity LLC, Aurora, CO) and a spacer at its N terminus, bears a K10C substitution mutation and has another cysteine residue (C57) added to the C-terminus. The two cysteine residues were labeled with Alexa Fluor 488 and Alexa Fluor 594. Details of the expression, purification, and dye-labeling of biotinylated GB1 have been described (*9*).

Synthetic DNA encoding 37 amino acids of the FBP WW domain was expressed in fusion with the 6His-GB1 domain at its N-terminus separated by a Fxa protease cleavage site. The WW domain contains a substitution mutation W30A, to prevent fibril formation (*8*) and cysteine residues flanking its termini. Similar in strategy to the GB1 construct, it is also flanked at its N-terminus by the AviTag-spacer sequence (Fig. S1). Thus, the DNA insert encoding 6His-GB1-Fxa-AviTag-Spacer-WW was cloned between the Nco1 and BamH1 sites of pET15b vector (Novagen, San Diego, CA). The resulting construct (6H-GB1-Avi-WW$^{W30A/CC}$) was verified by DNA sequencing.

The expression construct 6H-GB1-Avi-WW$^{W30A/CC}$ and a plasmid with an isopropylthiogalactoside (IPTG) inducible birA gene to over-express the biotin ligase (Avidity LLC) were co-transformed into *E.coli* BL-21 (DE3; Stratagene, La Jolla, CA). Cells were grown in Luria-Bertani medium, and expression was induced at an absorbance of 0.7 monitored at 600 nm with a final concentration of 1 mM IPTG for a period of 3-4 h. A final concentration of 50 µM d-biotin (Sigma, St. Louis, MO) was added to the medium ~ 30 min before induction. Typically, cells harvested from a 500-mL culture were lysed by uniform suspension in 90 mL of bacterial protein extraction reagent (B-PER, Pierce, Rockford, IL) containing 5 mM benzamidine and sonication. The lysate was centrifuged at 12,800 rpm (SS-34 rotor, ThermoFisher Scientific, Asheville, NC) for 30 min at 4ºC. The supernatant was subjected to affinity chromatography using streptavidin Mutein matrix (Roche Diagnostics GmbH, Mannheim, Germany). The column was equilibrated and washed extensively, after passing the lysate, with 1X PBS (1.7 mM KH2PO4, 5 mM Na2HPO4, 150 mM NaCl, pH 7.4) and the biotinylated 6H-GB1-Avi-WW$^{W30A/CC}$ was eluted in 1X PBS containing 2 mM d-biotin. Peak fractions were pooled, dialyzed against 25 mM Tris-HCl at pH 8, 100 mM NaCl, 2 mM CaCl$_2$ and 20 mM imidazole and subjected to protease cleavage at a concentration of 1-2 mg/ml fusion protein and 0.1-0.2% Fxa overnight at room temperature. The extent of cleavage was monitored by SDS-PAGE and the cleaved Avi-WW$^{W30A/CC}$ was attained in the flow-through by passing the digest on a Ni-NTA agarose affinity column. The flow-through fraction was incubated with 5 mM dithiothreitol for 1 hr at room temperature and further purified on reverse-phase HPLC (POROS 20 R2 resin, Perseptive Biosystems, Framington, MA) by eluting using a linear gradient from 99.95% water (v/v) and 0.05% TFA to 60% acetonitrile (v/v), 0.05% TFA (v/v) and 39.95% water (v/v) over a

period of 16 min at a flow rate of 4 mL/min. Aliquots of the peak fraction were lyophilized and stored at -70ºC. Biotin ligation to the biotin acceptor peptide was confirmed by mass spectrometry. An observed mass of 7717 was clearly indicative of the combined mass of biotin (expected mass of 226) ligated to the acceptor sequence in Avi-WW$^{W30A/CC}$ (expected mass of 7491). Labeling of Alexa Fluor 488 and Alexa Fluor 647 and purification of Avi-WW$^{W30A/CC}$ containing the donor-acceptor pair by size-exclusion (Superdex-30, 1.6 cm x 60 cm, GE HealthCare) followed by anion-exchange chromatography were performed as described previously with only slight modifications (*9*).

A

AviTag　　　　　　**Linker**

GMSGLNDIFE AQ**K**IEWHESS GLVPRGSH

MQYKLILNG**C** TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK TFTVTE**C**

GB1

B

AviTag　　　　　　**Linker**

GHMGMSGLND IFEAQ**K**IEWH ESSGLVAGG

CGATAVSEWT EYKTADGKTY YYNNRTLEST AEKPQELK**C**

FBP28 WW domain
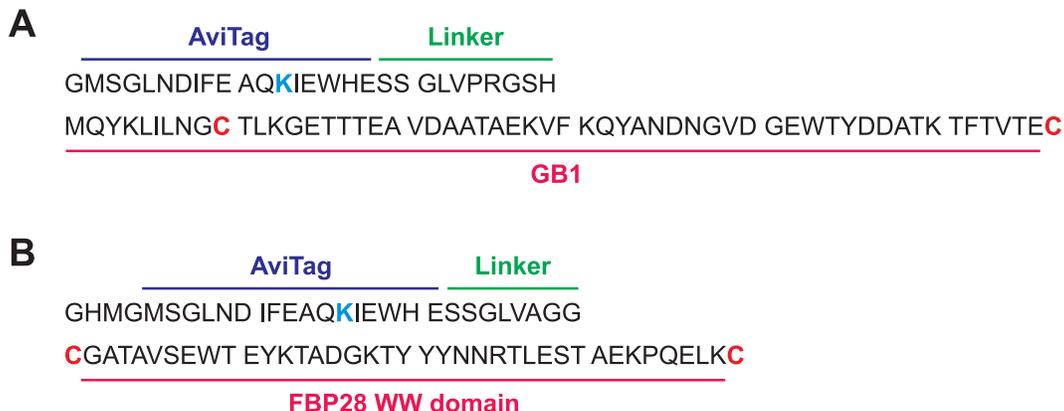
Fig. S1. Amino acid sequences of polypeptides containing protein GB1 (**A**) and the FBP WW domain (**B**). Dyes were attached to the cysteine residues (red) and a biotin molecule to the lysine residue (blue) in AviTag sequence.

Single molecule spectroscopy. Single molecule FRET experiments were performed using a confocal microscope system (MicroTime200, Picoquant). The CW mode of a dual mode (CW/pulsed) 485 nm diode laser (LDH-D-C-485, PicoQuant) was used to excite donor dyes (Alexa Fluor 488) through an oil-immersion objective (PlanApo, NA 1.4, × 100, Olympus). Donor and acceptor (Alexa Fluor 594 or Alexa Fluor 647) fluorescence was collected by the same objective, split into two channels, and focused through optical filters (ET525/50m for the donor and E600LP for the acceptor, Chroma Technology) onto photon-counting avalanche photodiodes (SPCM-AQR-15, PerkinElmer Optoelectronics). Additional details for the optical setup and single molecule experiments can be found elsewhere (*9, 27*).

Protein molecules were immobilized on a biotin-embedded, polyethyleneglycol (PEG)-coated glass coverslip (Bio_01, Microsurfaces Inc.) via a biotin (surface)-streptavidin-biotin (protein) linkage. To reduce photobleaching and populating triplet states of the dyes, a combination of chemicals were added to the 50 mM phosphate or HEPES buffer (pH 7.6) solution. For the protein GB1 experiments, (S)-trolox methyl ether (93510, Sigma) and 10 mM cysteamine (30070, Sigma) were added (*28*). For the WW domain experiments in 2 M GdmCl and 2 M GdmCl/24% glycerol solutions, 1 mM L-ascorbic acid (A92902, Sigma) and 1 mM methyl viologen (856117, Sigma) were added (*29*). For the experiment in 50% glycerol, 10 mM ascorbic acid and 10 mM methyl viologen were used.

To collect a large number of trajectories, we used an automated data collection scheme. An area of 10×10 μm^2 was raster scanned at low intensity (0.2 kW/cm^2) and the location of

molecules was determined. To ensure the single molecule detection, molecules immobilized too close to one another were identified by an image larger than a threshold size and were excluded. The piezo-controlled stage was then moved to locate each molecule and the trajectory was collected at high illumination intensity. The laser was turned off during movement of the stage to prevent photobleaching.. After the collection of trajectories for all identified molecules was completed, the procedure was repeated for the next 10×10 μm^2 area (49 areas in total). Before raster scanning each area, the focus along z-axis (perpendicular to the surface) was set at the position with minimum variance of the reflected image from the surface recorded by a CCD camera. The illumination intensity for the raster scan (0.2 kW/cm^2) and for the trajectory collection (2 or 20 kW/cm^2) were adjusted by inserting a neutral density filter (OD1 or OD2) in the laser path using a home-built mechanical shutter.

Measurement of relative viscosities. The relative viscosities of the solutions were obtained from the ratio of the translational diffusion times (proportional to the solvent viscosity) measured in fluorescence correlation spectroscopy (FCS) experiments on a rigid rod molecule, 20-residue poly-proline (*30*). Glycine and cysteine residues were added to the N- and C-termini of the peptide, and were labeled with Alexa Fluor 594 succinimidyl ester and Alexa Fluor 488 maleimide dyes, respectively.

Fig. S2 shows the donor-acceptor cross-correlation as a function of time delay (τ) for poly-proline 20 in 2 M GdmCl, 2 M GdmCl/24% glycerol, and 3 M GdmCl/50% glycerol. The three decay curves were fitted with a 3-dimensional diffusion equation (*31*)

$$C(\tau) \;=\; \frac{1}{N}\left(1 + \tau/\tau_D\right)^{-1}\left(1 + \alpha^{-2}\tau/\tau_D\right)^{-1/2} \;, \tag{S1}$$

where τ_D is the translational diffusion time, α is a constant determined by the ratio of the depth of focus and the beam waist at the focus, and N is the average number of molecules in the focal volume. We assumed that α is the same for all solutions. The fitted curves (dashed lines) shown in Fig. S2 used $\alpha^2 = 10$. To estimate an accurate diffusion time, α should be measured by an independent experiment for a species with a known diffusion time but the relative values (ratio of the diffusion times) are insensitive to α.
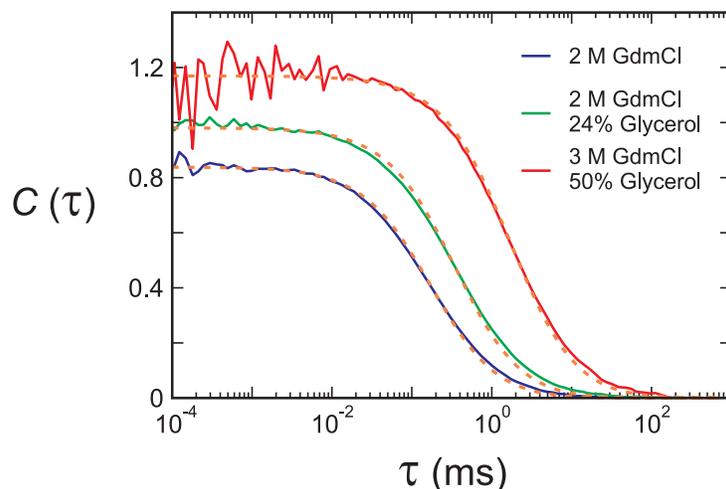
Fig. S2. Measurement of relative viscosities of solutions using donor-acceptor cross-correlation of Alexa Fluor 488 and Alexa Fluor 594 labeled poly-proline 20 in an FCS experiment. The correlation function $C(\tau)$ is plotted versus the time delay τ and fitted with Eq. S1 to obtain the diffusion times. The relative viscosities calculated from the ratio of the diffusion times are summarized in Table S1.

Kinetics of WW domain. Trajectories for the WW domain were collected at low intensity (2 kW/cm^2, ~ 70 photons/ms count rate) to obtain the folded populations and rate coefficients. As shown in Fig. S3A, the kinetics in 2 M guanidinium chloride (GdmCl), close to the expected denaturation midpoint, are so fast that waiting times and transitions between folded and unfolded states are not resolved when photons are collected in 1 ms bins. In this case, it is not possible to construct FRET efficiency trajectories to locate the transition points. This finding of sub-millisecond kinetics is consistent with temperature jump studies by Fersht and coworkers in which the relaxation time at the midpoint denaturant concentration is ~ 250 μs at 10°C (*8*). Additional evidence for sub-millisecond kinetics is apparent in the FRET efficiency distribution in Fig. S3A, which shows a single peak, indicating averaging of the FRET efficiencies for the folded and unfolded molecules due to multiple transitions within the 1 ms bin time, a phenomenon similar to fast chemical exchange in NMR (*10, 32, 33*).

Using the maximum likelihood method, however, the photon trajectories could as before (*33*) be analyzed directly without time binning to yield the FRET efficiencies of the folded and unfolded states, the relative population of the folded state, and the sum of the folding and unfolding rate coefficients (Table 1), $k = 15$ ms^{-1}. Under these conditions, there are on average less than 10 photons emitted from each of the folded and unfolded state segments of a trajectory (photon count rate is less than 100 ms^{-1}). Although there is a signature of the alternating folded and unfolded states in the photon trajectory in Fig. S3A, the photon count rate is insufficient to resolve the folded and unfolded states with high precision. The situation is similar in 2 M GdmCl/24% glycerol though a slightly broader FRET efficiency distribution indicates a slower inter-conversion between the folded and unfolded states as shown in Fig. 3B.

As shown in Fig. 3C, transitions between states become resolvable in the binned and photon trajectories in the 10-times more viscous solution and the FRET efficiency distribution becomes significantly broader. The parameters extracted from the maximum likelihood analysis of the photon trajectories using a two-state model are summarized in Table S1.

5
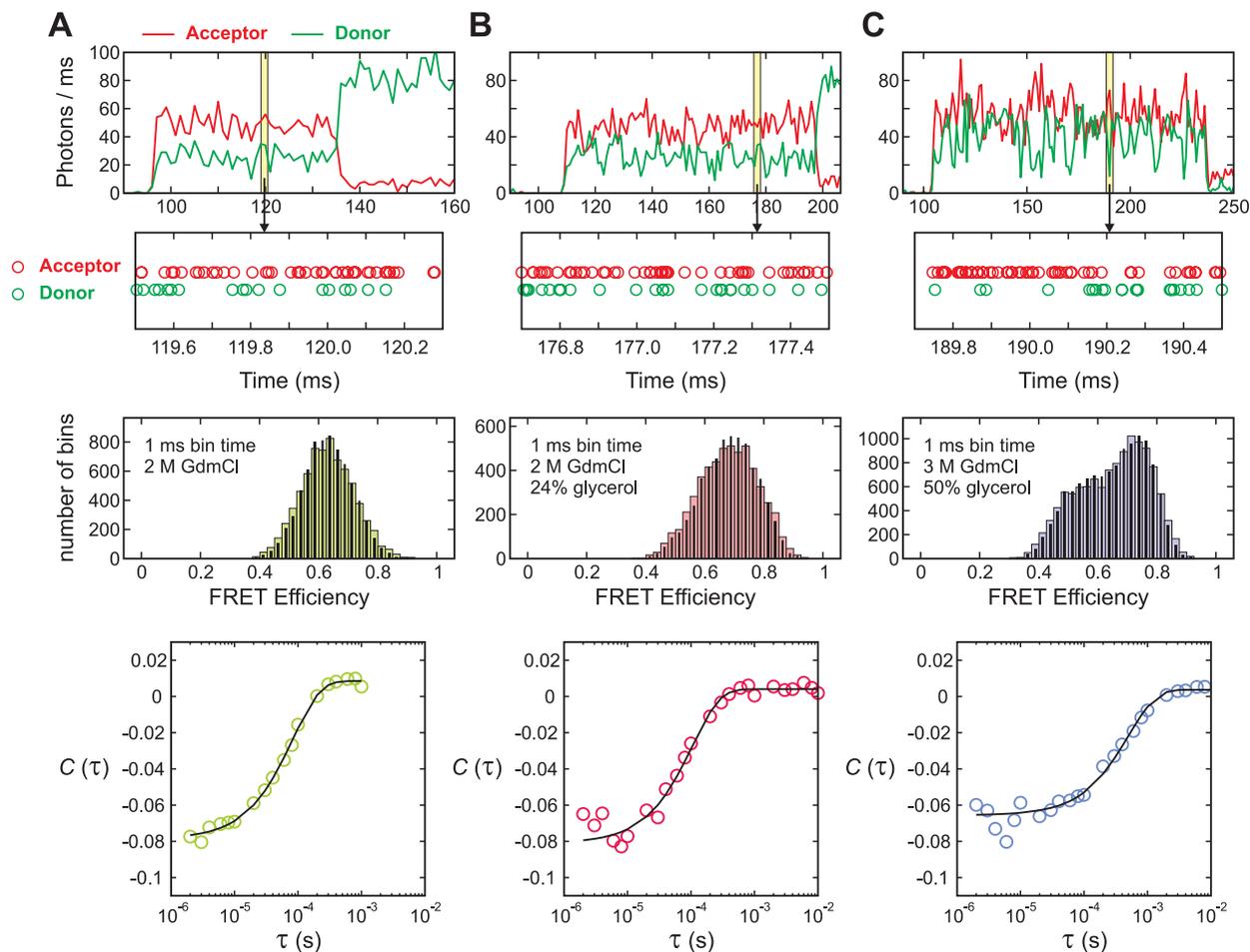
Fig. S3. Representative binned fluorescence trajectories of donor (green) and acceptor (red), photon trajectories, and FRET histograms of WW domain at various experimental conditions. Trajectories of WW domain were collected at low illumination intensity (2 kW/cm^2) in 2 M GdmCl (**A**), 2 M GdmCl/24% glycerol (**B**), and 3 M GdmCl/50% glycerol (**C**). The slowing down of folding and unfolding kinetics with increasing glycerol/water ratio is apparent from the larger fluctuations in the binned trajectories (1 ms bin time, top) and the broader FRET efficiency distribution (3rd row). Strings of arrival times of donor and acceptor photons (photon trajectory, middle) in the yellow shaded regions in top panels show that folding and unfolding gets slower with the increasing glycerol content. The FRET efficiency histograms were constructed from the trajectories with the mean photon count rate > 50 ms^{-1}. The histograms constructed from re-colored photon trajectories (black narrow bars) using the parameters obtained from the maximum likelihood method coincide extremely well with the experimental histograms (colored wide bars), indicating that the model provides an excellent description of the data (*10*). (bottom) Donor-acceptor cross correlation functions were calculated to compare the relaxation rate with the sum of folding and unfolding rate coefficients obtained from the maximum likelihood method. The parameters are listed in Table S1.

6

Viscosity dependence of average transition-path time. To obtain the solvent viscosity (η) dependence of the average transition-path time from the viscosity dependence of the folding and unfolding rates, we must determine how much glycerol changes these rates from its effect on stability. To this end we employed a linear free energy relation, i.e. $k_F \propto K^{\beta}$ and $k_U \propto K^{\beta-1}$, where $0 \leq \beta \leq 1$. An additional factor to consider is that the scaling with viscosity may not show the simple $1/\eta$ dependence predicted from Kramers theory. Including the possibility of a contribution from internal friction, we used the relation introduced by Ansari *et al.* (*34*) where the total viscosity is the sum of an internal viscosity, σ, and the solvent viscosity to give $k_F \propto (\sigma + \eta)^{-1}K^{\beta}$ and $k_U \propto (\sigma + \eta)^{-1}K^{\beta-1}$. The data in Table S1 show that σ is negligible and $\beta \approx 1$, as indicated by the approximate scalings: $k_F \propto \eta^{-1}K$ and $k_U \propto \eta^{-1}$. Therefore, both rate coefficients are proportional to $1/\eta$, and therefore also the average transition-path time.

Table S1. Viscosity dependence of kinetics parameters of WW domain*

	2 M GdmCl	2 M GdmCl 24% Glycerol	3 M GdmCl 50% Glycerol
E_F	0.839 (±0.002)	0.826 (±0.002)	0.766 (±0.001)
E_U	0.484 (±0.002)	0.489 (±0.003)	0.481 (±0.001)
k (ms^{-1}) Maximum Likelihood	15.5 (±0.4)	9.87 (±0.28)	2.42 (±0.05)
DA cross correlation	12.1 (±0.5)	9.50 (±0.89)	2.00 (±0.10)
p_F	0.415 (±0.006)	0.567 (±0.006)	0.596 (±0.005)
K (= $p_F/(1- p_F)$)	0.709	1.31	1.48
k_F (= $k \cdot p_F$) ms^{-1}	6.42	5.60	1.44
k_U (= $k \cdot (1-p_F)$) ms^{-1}	9.05	4.27	0.98
Relative solvent viscosity	1	2.0	9.8

* Errors are standard deviations obtained from the curvature at the maximum of the likelihood function (maximum likelihood method) or chi-square function (donor-acceptor cross correlation calculation).

Calculation of likelihood function for one-step model. The average transition-path time, $\langle t_{TP} \rangle$, or its upper bound can be determined using the calculation of the likelihood function for a model with a virtual intermediate state S, and identifying the lifetime of this state, τ_S, with $\langle t_{TP} \rangle$. A likelihood function is a relative probability factor that a given photon trajectory is observed, which can be calculated using parameters of a kinetic model. Fig. S4A shows a simple photon trajectory containing a single folding transition. For a two-state model with a finite transition-path time, each photon is emitted from one of the three states F, S, and U. Therefore, one can label each photon with state indices to generate state trajectories. Since there is only one folding transition, state trajectories consist of a segment of U followed by a segment of S and a segment

of F. m and n are the photon indices of the last and the first photon of the folded and unfolded segments, respectively. The likelihood function $L(\tau_S)$ of this photon trajectory is the summation of the likelihood functions of the state trajectories $L(m,n,\tau_S)$. However, calculating $L(m,n,\tau_S)$ for individual state trajectories is impossible because there are too many state trajectories (3^{N-2} trajectories for N photons). We used a method previously developed by Gopich and Szabo (*10*) as described below.
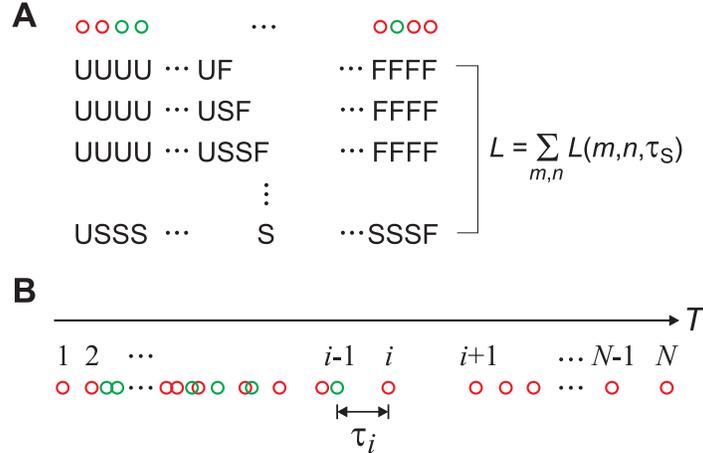
Fig. S4. (**A**) Photon trajectory with a single transition. Red and green circles indicate acceptor and donor photons, respectively. Labeling photons with state indices generates state trajectories. Here, the likelihood for a state trajectory is a function of the indices of the transition from U to S (*m*) and S to F (*n*), and τ_S. The likelihood of the photon trajectory is the summation of these likelihoods of the states trajectories. (**B**) The definition of photon indices and time interval used in the calculation of the likelihood function in Eq. S9.

In a two-state model with an instantaneous transition, the equilibrium and kinetic parameters are given by

$$p_F + p_U = 1$$
$$k = k_F + k_U \quad , \qquad\qquad\qquad \text{(S2)}$$
$$k_F p_U = k_U p_F$$

where p_F and p_U are the fraction of the folded and unfolded states, and k_F and k_U are folding and unfolding rate coefficients. The rate matrix is

$$\mathbf{K} = \begin{pmatrix} -k_U & k_F \\ k_U & -k_F \end{pmatrix}. \qquad\qquad\qquad \text{(S3)}$$

This was the model used in calculating the parameters in Table S1. In a model with a finite transition-path time (see Fig. 4A),

$$k_{U'} p_{F'} = k_S p_S = k_{F'} p_{U'} \qquad\qquad\qquad \text{(S4)}$$

$$p_{F'} + p_{U'} + p_S = 1 . \tag{S5}$$

$$\mathbf{K} = \begin{pmatrix} -k_{U'} & k_S & 0 \\ k_{U'} & -2k_S & k_{F'} \\ 0 & k_S & -k_{F'} \end{pmatrix} . \tag{S6}$$

Here, notations F' and U' are used for the folded and unfolded states to distinguish these from those in the two state model with an instantaneous transition. Since the sum of the average time that a molecule spends in the unfolded state and a half of the time spent in S in the finite transition-path time model should be the same as the time spent in the unfolded state in the model with an instantaneous transition for a long trajectory with multiple transitions ,

$$\frac{1}{k_F} = \frac{1}{2}\left(\frac{1}{k_{F'}} + \frac{1}{2k_S}\right) + \frac{2}{2^2}\left(\frac{1}{k_{F'}} + \frac{1}{2k_S}\right) + \frac{3}{2^3}\left(\frac{1}{k_{F'}} + \frac{1}{2k_S}\right) + \cdots = 2\left(\frac{1}{k_{F'}} + \frac{1}{2k_S}\right) \approx \frac{2}{k_{F'}}, \tag{S7}$$

where each term corresponds to a trajectory with 0, 1, 2, ….transitions from U to S and back to U. Similarly,

$$\frac{1}{k_U} = 2\left(\frac{1}{k_{U'}} + \frac{1}{2k_S}\right) \approx \frac{2}{k_{U'}} . \tag{S8}$$

In this analysis we assume that $\tau_S = 1/2k_S$ is also $\langle t_{TP} \rangle$.

The likelihood function of this model for the jth photon trajectory is (*10*)

$$L_j = \mathbf{v}_{fin}^{\text{T}} \prod_{i=2}^{N} \{\mathbf{nF}(c_i) \exp[(\mathbf{K} - \mathbf{n})\tau_i]\} \, \mathbf{nF}(c_1)\mathbf{v}_{ini} \tag{S9-1}$$

$$L_j = \mathbf{v}_{fin}^{\text{T}} \prod_{i=2}^{N} [\mathbf{F}(c_i) \exp(\mathbf{K}\tau_i)] \, \mathbf{F}(c_1)\mathbf{v}_{ini} , \tag{S9-2}$$

where N is the number of photons in a trajectory, c_i is the color of the ith photon (donor or acceptor), and τ_i is a time interval between the ith and $(i\text{-}1)^{\text{th}}$ photons as shown in Fig. S4. The photon color matrix $\mathbf{F}$ depends on the color of a photon as $\mathbf{F}(acceptor) = E$ and $\mathbf{F}(donor) = \mathbf{I} - E$, where E is a diagonal matrix with elements that are FRET efficiencies of the individual states. $\mathbf{n}$ is a diagonal matrix with elements that are photon count rates of the individual states. $\mathbf{v}_{ini}$ and $\mathbf{v}_{fin}$ are state vectors at the beginning and the end of the trajectory. For the case of a folding transition in Fig. S4A, $\mathbf{v}_{ini} = (0\ 0\ 1)^{\text{T}}$ and $\mathbf{v}_{fin} = (1\ 0\ 0)^{\text{T}}$. We used Eq. S9-1 for protein GB1 to utilize the significantly different photon count rates in the folded ($n_F = 250$ ms^{-1}) and unfolded states ($n_U = 460$ ms^{-1}) resulting from fluorescence quenching in the folded state (*9*). On the other hand, a reduced form, Eq. S9-2, for the case of $n_F = n_U$ was used for WW domain with nearly equal photon count rates in both states, which considerably simplifies the analysis. The calculation of likelihood values were performed using the diagonaliztion of the matrix exponential in Eq. S9 as described in ref (*10*). Practically, the log likelihood function was calculated and the total log

likelihood function of all trajectories were calculated by the summation of individual log likelihood functions as $\ln L = \sum_j \ln L_j$.

For protein GB1, as described in ref. (*9*), a photon trajectory was initially split into the folded and unfolded segments by finding the photon interval with the maximum transition probability (transition interval) using an instantaneous transition model. To calculate the likelihood function of an individual trajectory, we used E_F, E_U, $E_S = (E_F + E_U)/2$, n_F, n_U, $n_S = (n_F + n_U)/2$ values calculated from that trajectory as there are variations in the FRET efficiency due to the inhomogeneity in the environment (*35*). We used the previously measured folding and unfolding rate coefficients of protein GB1 $k_F = 1.12$ s^{-1} and $k_U = 0.41$ s^{-1} (*9*).

For the WW domain, k_F and k_U were obtained from the low intensity measurement as listed in Table S1. The folded and unfolded segments were assigned using the Viterbi algorithm (*36, 37*), adapted for photon trajectories (*33*), and these rate coefficients as described in ref. (*33*). When there are multiple transitions in a trajectory, each pair of segments with a single folding or unfolding transition was analyzed separately. In the calculation of the likelihood function in Eq. S9-2 for these segments, we reduced the rate coefficients by a factor of 1000 to effectively eliminates the possibility of multiple transitions, i.e. k$_{U'}$ (= k_U/500) and k$_{F'}$ (= k_F/5000) (Eq. S7). This treatment is valid since we use the difference of the likelihood values (ΔL) in this paper.
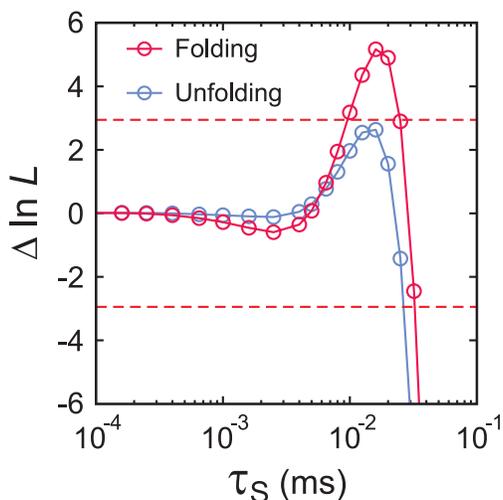
Fig. S5. The difference of the log likelihood, $\Delta \ln L = \ln L(\tau_S) - \ln L(0)$, plotted as a function of τ_S for the folding and unfolding transition data of WW domain in 3 M GdmCl/50% glycerol at high illumination intensity. Transition-path times for folding (17.1 ($\pm$3.6) μs) and unfolding (14.6 ($\pm$3.6) μs) are the same within experimental error as required by theory (*12*) and by our model.

Selection of trajectories without photophysical artifacts. Trajectories were selected for analysis that were free of photophysical artifacts using the following criteria. First, as described above, a photon trajectory with a single transition was split into two pieces belonging to the folded and unfolded states and the FRET efficiencies of the two states were calculated for each trajectory. If one or both of these values of the FRET efficiency deviated from the mean value for all trajectories by more than 2 standard deviations (95%), the trajectory was discarded. Using

this criterion ~ 10% of the trajectories were discarded. As shown previously from polarization meaurements, about 5% of trajectories for protein GB1 showed abnormal FRET efficiency segments that arose from sticking of the unfolded protein to the surface altering the FRET efficiency because of a lack of incomplete κ^2 averaging (*9, 35*).

Trajectories containing acceptor blinking near the transition were not included in the analysis. In a photon trajectory, a string of photons was assigned to acceptor blinking when the FRET efficiency of this string (E_s) was too low to belong to the unfolded states (E_U), namely, $E_U - E_s < 2(E_U(1-E_U)/N_s)^{1/2}$, where N_s is the number of photons in the string which contains at least 8 consecutive donor photons. If this acceptor blinking region was within the ±30 μs window of the transition interval, the trajectory was discarded (8 discarded trajectories for protein GB1 and 30 for WW domain).

Finally, 10 protein GB1 trajectories were discarded because of abnormally large fluctuations in the count rate from donor blinking that resulted in mis-assignment of the most probable transition interval.

Simulation of photon trajectories for WW domain. To demonstrate the validity of the method for determining $\langle t_{TP} \rangle$ for the WW domain, photon trajectories were simulated with experimental parameters and various values of τ_S, and analyzed in exactly the same way as done for the experimental data.

The details of the method for generating photon trajectories using given FRET efficiencies, photon count rates, and rate coefficients have been described in ref. (*33*). The following experimental parameters were used in the present calculation: $E_F = 0.8$, $E_U = 0.49$, $n = 650$ photons/ms, $\sigma_n = 150$ photons/ms, $L_{trj} = 0.37$ ms, and $\sigma_{trj} = 0.17$ ms. For each trajectory, the photon count rate and the length of the trajectory were varied from n and L_{trj} by their standard deviations σ_n and σ_{trj}. The trajectories were accepted when only one transition occured from the folded to the unfolded state or from the unfolded to the folded state, only one transition was detected by the Viterbi algorithm, and the lengths of both the folded and unfolded segments were longer than 50 μs. A total of 5 sets of ~ 500 photon trajectories were simulated for each value of τ_S of 0.5, 2, 5, 10, 20, 50 μs. The calculation of the likelihood function for each ~ 500-photon trajectory set are plotted in Fig. S6.

Both the appearance of a peak for the shortest τ_S and the height of the peak vary due to the statistical fluctuations associated with a finite number of photons. For τ_S of 0.5 - 5 μs most of the likelihood maxima are below the 95% confidence level, so 500 transitions and $n = 650$ photons/ms are insufficient to determine $\langle t_{TP} \rangle$. Only an estimation of an upper bound for τ_S and therefore $\langle t_{TP} \rangle$ is possible under these conditions. However, a $\tau_S > 10$ μs can be measured with high confidence. Interestingly, when the input τ_S is 50 μs, the τ_S value at the maximum is only 20 μs. This underestimation of τ_S results from the fact that the length of the trajectory is not long compared to τ_S. Since the average length is 0.37 ms, 14% of the photons belong to S. Therefore, the values of E_F and E_U that are used to calculate the likelihood function are lower and higher than true E_F and E_U, respectively. In this case, the likelihood value becomes smaller and the time at the maximum becomes shorter as well. Simulations with trajectories twice as long result in τ_S at the peak maximum at ~ 30 μs.
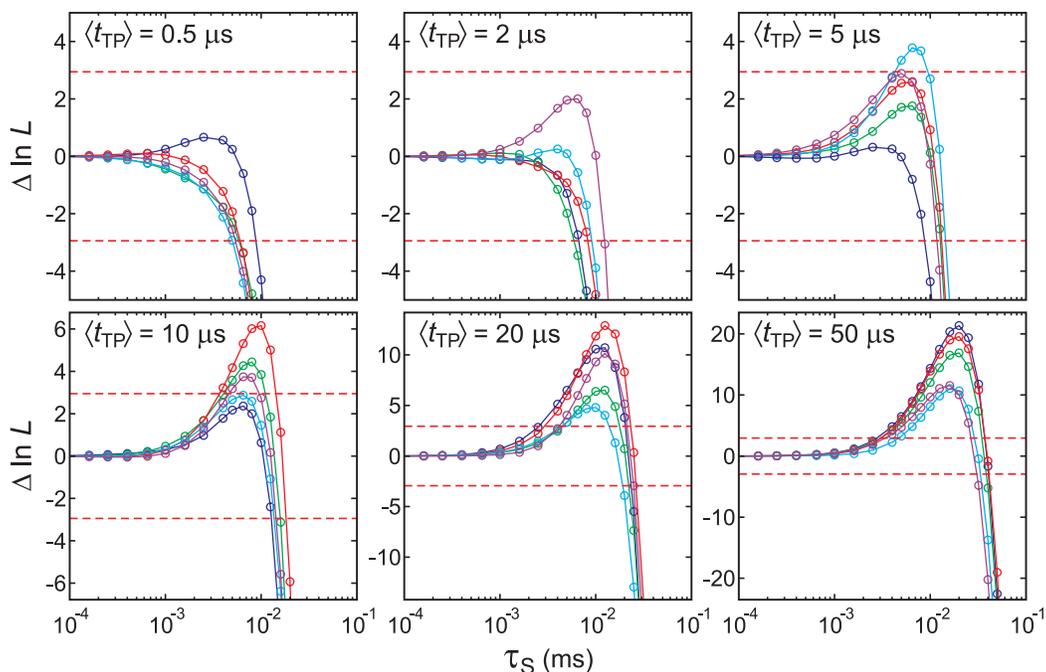
Fig. S6. Sum of the log likelihood functions of ~ 500 folding and unfolding transitions at various transition-path times. 5 data set were simulated for each transition-path time and plotted with different colors.

FRET efficiency of S. In the transition path analysis, we have assumed that the mean FRET efficiency E_S of the conformations in the transition path is the mean of those for the folded and unfolded states. To test this assumption, the likelihood function was calculated for varying FRET efficiency of S (Fig. S7A). Interestingly, a peak formed when E_S is between E_F of 0.8 and E_U of 0.5. For other values of E_S, the likelihood function decays very fast because the models with these FRET efficiencies are inconsistent with the data. We applied this method to one of the simulated data sets with $\tau_S = 20$ μs, which shows similar behavior as shown in Fig. S7B. This indicates that our assumption is valid although even the higher photon count rate and larger separation between E_F and E_U are required to actually assign an accurate mean FRET efficiency to the conformations in the transition path.
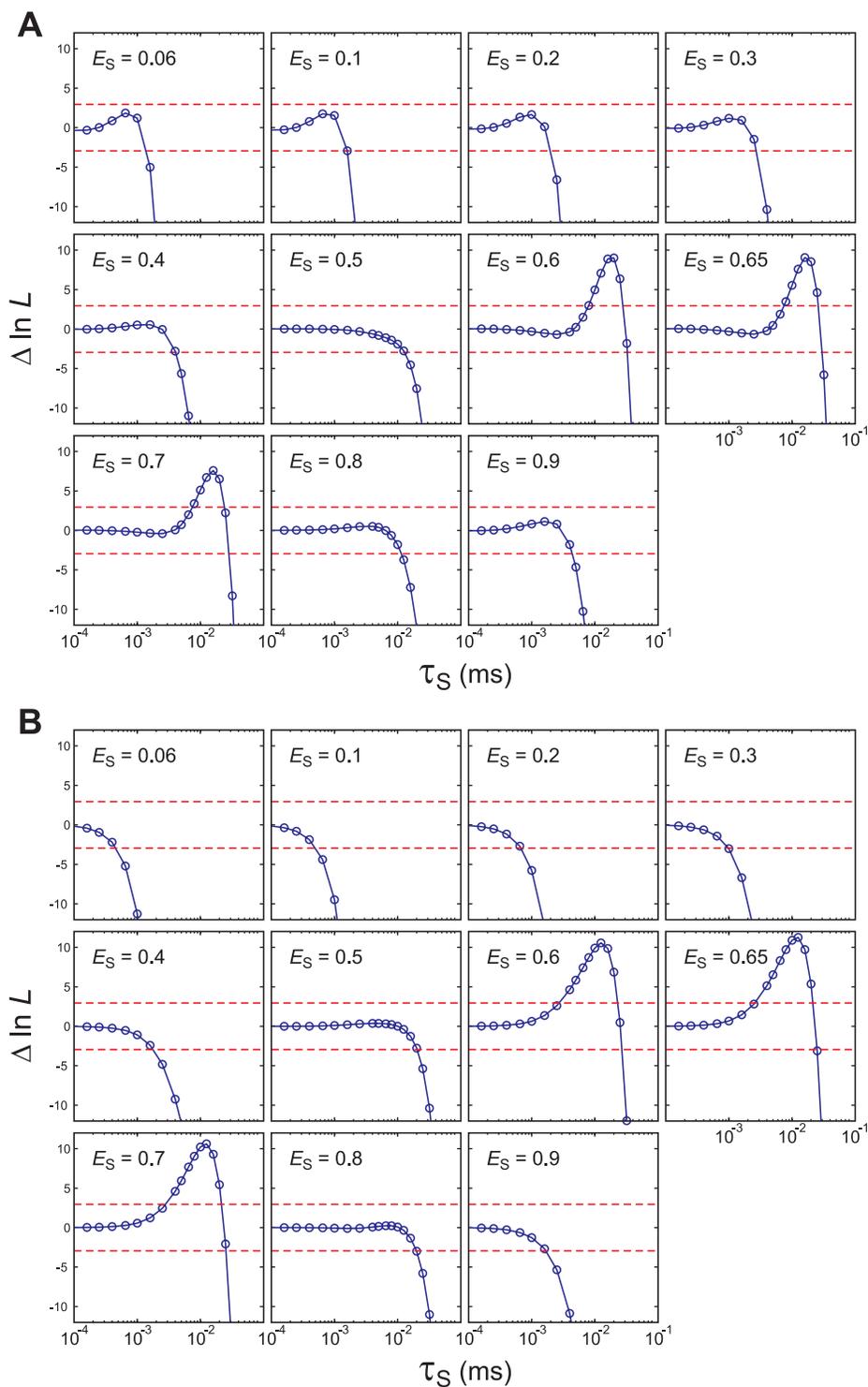
Fig. S7. Sum of the log likelihood functions of ~ 500 folding and unfolding transitions with various intermediate state FRET efficiency E_S values for the experimental data collected for WW domain in 3 M GdmCl/50% glycerol (**A**) and for one of the simulation data set with the $\langle t_{TP} \rangle$ = 20 μs (blue curve in Fig. S6) (**B**).

More realistic modeling of the transition path. In addition to a gradual increase in FRET efficiency along a transition path, a realistic distribution of transition-path times should show a peak rather than the monotonically decreasing exponential distribution of our one-step model (Fig. S8). A more gradual transition, as well as a more realistic distribution of transition-path times, can be realized by adding more steps to the model, represented by virtual intermediate states. When N states were introduced between the folded and unfolded states (Fig. S9A), the FRET efficiencies of these states were set to gradually increase from that of the unfolded state to that of the folded state as $E_{S_j} = [(N+1-j)E_F + jE_U]/(N+1)$ (Fig. S9B). We assumed that the rate coefficient escaping from the intermediate states to the nearby states are equal in each model (Fig. S9A). (k_S, k_{2S}, and k_{3S} for one, two, and three intermediate states models) The likelihood function in Eq. S9 for each model can be calculated simply by modifying the rate matrix **K** and using the relationship between the mean transition-path time and these rate coefficients. Below are derived the distributions of transition-path times and the mean transition-path times for the cases of 1 − 3 intermediate states.

1. One intermediate state. ($U \leftrightarrow S \leftrightarrow F$)

The rate matrix is

$$\mathbf{K} = \begin{pmatrix} -k_{U'} & k_S & 0 \\ k_{U'} & -2k_S & k_{F'} \\ 0 & k_S & -k_{F'} \end{pmatrix}. \tag{S10}$$

The distribution of transition-path times is then:

$$P(\tau) = 2k_S e^{-2k_S \tau}, \tag{S11}$$

and the average transition-path time $\langle t_{TP} \rangle = \langle \tau \rangle = 1/2k_S$.

2. Two intermediate states. ($U \leftrightarrow S_2 \leftrightarrow S_1 \leftrightarrow F$)

The rate matrix is

$$\mathbf{K} = \begin{pmatrix} -k_{U'} & k_{2S} & 0 & 0 \\ k_{U'} & -2k_{2S} & k_{2S} & 0 \\ 0 & k_{2S} & -2k_{2S} & k_{F'} \\ 0 & 0 & k_{2S} & -k_{F'} \end{pmatrix}. \tag{S12}$$

When, a molecule enters the transition path (U → S_2), the paths to make a successful transition to F without returning to U consists of oscillations between S_2 and S_1 followed by a final transition from S_1 to F as

14

$$\text{Path 1:} \quad S_2 \rightarrow S_1 \rightarrow F$$
$$\text{Path 2:} \quad S_2 \rightarrow S_1 \rightarrow S_2 \rightarrow S_1 \rightarrow F \tag{S13}$$
$$\text{Path 3:} \quad S_2 \rightarrow S_1 \rightarrow S_2 \rightarrow S_1 \rightarrow S_2 \rightarrow S_1 \rightarrow F$$
$$\vdots$$

For paths 1 and 2, the distributions of transition-path times are

$$P_1(\tau) = \int_0^\tau (2k_{2S})^2 e^{-2k_{2S}t_1} e^{-2k_{2S}(\tau-t_1)} dt_1 = (2k_{2S})^2 \tau e^{-2k_{2S}\tau} \tag{S14}$$

$$P_2(\tau) = \int_0^\tau \int_0^{t_1} \int_0^{t_2} (2k_{2S})^4 e^{-2k_{2S}t_3} e^{-2k_{2S}t_2} e^{-2k_{2S}t_1} e^{-2k_{2S}(\tau-t_1-t_2-t_3)} dt_3 dt_2 dt_1$$
$$= (3!)^{-1} (2k_{2S})^4 \tau^3 e^{-2k_{2S}\tau} \tag{S15}$$

In general, the distribution of transition-path times for path n is

$$P_n(\tau) = \frac{1}{(2n-1)!} (2k_{2S})^{2n} \tau^{2n-1} e^{-2k_{2S}\tau} . \tag{S16}$$

The total distribution of the transition-path times is the sum of the individual transition-path time distributions in Eq. S16 weighted by the relative probability of the paths. Since the probabilities of forward and backward transitions are the same, a factor 1/2 should be multiplied for each transition made. Then, the total transition-path time distribution becomes

$$P(\tau) = \sum_{n=1}^{\infty} 2^{-2n} P_n(\tau) \Big/ \sum_{n=1}^{\infty} 2^{-2n} = 3k_{2S} e^{-2k_{2S}\tau} \sinh k_{2S}\tau . \tag{S17}$$

The average transition-path time can be calculated as

$$\langle t_{TP} \rangle = \langle \tau \rangle = \int_0^\infty \tau P(\tau) d\tau = \frac{4}{3k_{2S}} . \tag{S18}$$

3. Three intermediate states. ($U \leftrightarrow S_3 \leftrightarrow S_2 \leftrightarrow S_1 \leftrightarrow F$)

The rate matrix is:
$$\mathbf{K} = \begin{pmatrix} -k_{U'} & k_{3S} & 0 & 0 & 0 \\ k_{U'} & -2k_{3S} & k_{3S} & 0 & 0 \\ 0 & k_{3S} & -2k_{3S} & k_{3S} & 0 \\ 0 & 0 & k_{3S} & -2k_{3S} & k_{F'} \\ 0 & 0 & 0 & k_{3S} & -k_{F'} \end{pmatrix} . \tag{S19}$$

When, a molecule enters the transition path (U → S$_3$), the paths to make a successful trip to F without returning to U consist of collections of paths with the same number of transitions as

15

$$\begin{aligned}
\text{Class 1:} \quad & S_3 \to S_2 \to S_1 \to F \\
\text{Class 2:} \quad & S_3 \to S_2 \to S_1 \to S_2 \to S_1 \to F \\
& S_3 \to S_2 \to S_3 \to S_2 \to S_1 \to F \\
\text{Class 3:} \quad & S_3 \to S_2 \to S_1 \to S_2 \to S_1 \to S_2 \to S_1 \to F \\
& S_3 \to S_2 \to S_1 \to S_2 \to S_3 \to S_2 \to S_1 \to F \\
& S_3 \to S_2 \to S_3 \to S_2 \to S_1 \to S_2 \to S_1 \to F \\
& S_3 \to S_2 \to S_3 \to S_2 \to S_3 \to S_2 \to S_1 \to F \\
& \qquad\qquad\qquad \vdots
\end{aligned} \tag{S20}$$

Similar to Eq. S16, the transition-path time distribution of class n becomes

$$P_n(\tau) = \frac{1}{(2n)!}(2k_{3S})^{2n+1}\tau^{2n}e^{-2k_{3S}\tau} . \tag{S21}$$

Since there are 2^{n-1} paths in class n and $2n+1$ transitions in each path, the weighting factor is $2^{-(2n+1)} \cdot 2^{n-1} = 2^{-(n+2)}$. Then, the total distribution of transition-path times is

$$P(\tau) = \sum_{n=1}^{\infty} 2^{-(n+2)} P_n(\tau) \Big/ \sum_{n=1}^{\infty} 2^{-(n+2)} = 4k_{3S}e^{-k_{3S}\tau}\sinh^2\frac{k_{3S}\tau}{\sqrt{2}} . \tag{S22}$$

Finally, the average transition-path time can be calculated as

$$\langle t_{TP} \rangle = \langle \tau \rangle = \int_0^{\infty} \tau P(\tau)d\tau = \frac{5}{2k_{3S}} . \tag{S23}$$

The three transition-path time distributions of the three models are compared in Fig. S8. As the number of intermediate state increases, the distribution becomes narrower and the peak moves closer to the mean value. The likelihood functions for the experimental data of WW domain in 3 M GdmCl/50% glycerol are compared in Fig. S9C, which shows similar result with a slightly increasing transition-path times in the models with more intermediate states.
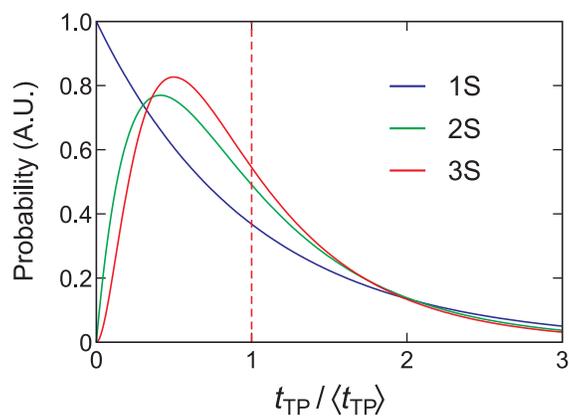
Fig. S8. Distribution of transition-path times in models with 1 – 3 intermediate states calculated using Eqs. S11, S17, and S22.
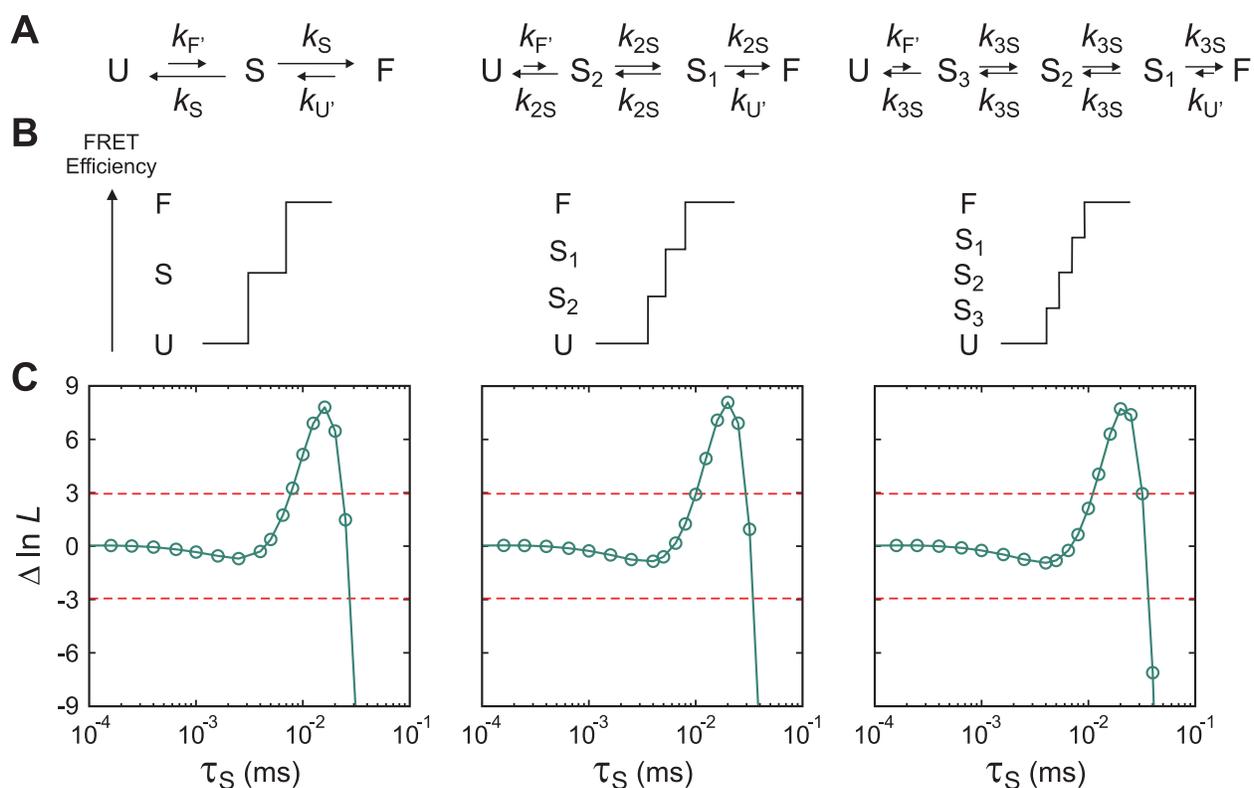
Fig. S9. Comparison of the transition path models. (**A**) Kinetics models. (**B**) The FRET efficiencies of the intermediate states increase from E_U to E_F equally. (**C**) Likelihood functions for WW domain experiment in 50% glycerol.

References

1. J. D. Bryngelson, J. N. Onuchic, N. D. Socci, P. G. Wolynes, *Proteins-Struct. Funct. Gen.* **21**, 167 (1995).
2. F. Noe, C. Schutte, E. Vanden-Eijnden, L. Reich, T. R. Weikl, *Proc. Natl. Acad. Sci. USA* **106**, 19011 (2009).
3. G. R. Bowman, V. S. Pande, *Proc. Natl. Acad. Sci. USA* **107**, 10890 (2010).
4. D. E. Shaw et al., *Science* **330**, 341 (2010).
5. S. Piana, K. Lindorff-Larsen, D. E. Shaw, *Biophys. J.* **100**, L47 (2011).
6. K. Lindorff-Larsen, S. Piana, R. O. Dror, D. E. Shaw, *Science* **334**, 517 (2011).
7. H. Nguyen, M. Jager, A. Moretto, M. Gruebele, J. W. Kelly, *Proc. Natl. Acad. Sci. USA* **100**, 3948 (2003).
8. M. Petrovich, A. L. Jonsson, N. Ferguson, V. Daggett, A. R. Fersht, *J. Mol. Biol.* **360**, 865 (2006).
9. H. S. Chung, J. M. Louis, W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **106**, 11837 (2009).
10. I. V. Gopich, A. Szabo, *J. Phys. Chem. B* **113**, 10965 (2009).
11. SI, Information on materials and methods is available as supporting online material on *Science Online*.
12. G. Hummer, *J. Chem. Phys.* **120**, 516 (2004).
13. F. Liu, M. Nakaema, M. Gruebele, *J. Chem. Phys.* **131**, 195101 (2009).
14. I. C. Yeh, G. Hummer, *J. Phys. Chem. B* **108**, 15873 (2004).
15. N. D. Socci, J. N. Onuchic, P. G. Wolynes, *J. Chem. Phys.* **104**, 5860 (1996).
16. In the case of protein GB1 there is the possibility of a sparsely populated intermediate between the folded and unfolded states (17-20). In this study we have implicitly defined the transition-path time for both the WW domain and protein GB1 in terms of just the two deep minima of the folded and unfolded states.
17. S. H. Park, M. C. R. Shastry, H. Roder, *Nat. Struct. Biol.* **6**, 943 (1999).
18. E. L. McCallister, E. Alm, D. Baker, *Nat. Struct. Biol.* **7**, 669 (2000).
19. A. Morrone et al., *Biophys. J.* **101**, 2053 (2011).
20. B. A. Krantz, L. Mayne, J. Rumbley, S. W. Englander, T. R. Sosnick, *J. Mol. Biol.* **324**, 359 (2002).
21. Clarke and coworkers (22) have found, for example, domains with similar structures and stability that have folding rates which differ by ~3,000-fold. The slower-folding domains show very little dependence on solvent viscosity, suggesting a large internal friction and therefore a much smaller D* (22, 23).
22. B. G. Wensley et al., *Nature* **463**, 685 (2010).
23. T. Cellmer, E. R. Henry, J. Hofrichter, W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **105**, 18320 (2008).
24. J. Kubelka, J. Hofrichter, W. A. Eaton, *Curr. Opin. Struct. Biol.* **14**, 76 (2004).
25. R. B. Best, G. Hummer, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6732 (2005).
26. J. Kubelka, E. R. Henry, T. Cellmer, J. Hofrichter, W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **105**, 18655 (2008).
27. K. A. Merchant, R. B. Best, J. M. Louis, I. V. Gopich, W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **104**, 1528 (2007).
28. L. A. Campos *et al.*, *Nature Methods* **8**, 143
29. J. Vogelsang *et al.*, *Angew.Chem.* **47**, 5465 (2008).
30. R. B. Best *et al.*, *Proc. Natl. Acad. Sci. USA* **104**, 18964 (2007).
31. S. R. Aragon, R. Pecora, *J. Chem. Phys.* **64**, 1791 (1976).
32. I. Gopich, A. Szabo, *J. Chem. Phys.* **122**, 014707 (2005).
33. H. S. Chung *et al.*, *J. Phys. Chem. A* **115**, 3642 (2011).
34. A. Ansari, C. M. Jones, E. R. Henry, J. Hofrichter, W. A. Eaton, *Science* **256**, 1796 (1992).
35. H. S. Chung, J. M. Louis, W. A. Eaton, *Biophys. J.* **98**, 696 (2010).
36. A. J. Viterbi, *IEEE Trans. Information Theory* **13**, 260 (1967).
37. L. R. Rabiner, *Proc. IEEE* **77**, 257 (1989).