Editors' Suggestion | Featured in Physics

Cooperative Native Contact Formation Facilitates Free Energy Barrier Crossing in Protein Folding

Chi-Jui Feng,[1] Ulrich Baxa,[2] John M. Louis,[1] and Hoi Sung Chung[1,*]
[1]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA
[2]Multi-Institute Cryo-EM Facility, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA

(Received 25 September 2025; accepted 22 January 2026; published 9 March 2026)

Understanding the folding mechanisms of biomolecules, such as proteins and nucleic acids, requires probing time-dependent changes of conformations, including secondary and tertiary structure formation. All these mechanism-relevant structural transitions, typically seen as coarse-grained descriptions in molecular dynamics simulations, occur in the tiny fraction of a molecular trajectory called the transition path (TP). However, TPs have been inaccessible for most proteins due to the limited time resolution of single-molecule techniques. Here, we measure the TP times of eight two-state single-domain proteins in aqueous solution using nanophotonics-enhanced single-molecule fluorescence spectroscopy. We found that the TP times are extremely short, ranging from 0.7 to 4 μs, and insensitive to most protein properties. Surprisingly, however, the extracted diffusion coefficient at the free energy barrier on the number of native contacts coordinate increases with protein length. This facilitation of diffusion in larger proteins is explained by increased cooperative native contact formation during folding, which reduces the roughness of the energy landscape. Our results reveal that evolutionary optimization of the energy landscape is much more efficient for protein folding than for other biomolecular processes.

DOI: 10.1103/2q9m-4scc

The transition path (TP) is part of a molecular trajectory connecting stable conformational states, such as folded and unfolded states for a two-state protein [Fig. 1(a)] [1–3]. Therefore, the TP encodes the dynamics of crossing a free energy barrier between these states. The TP can only be observed by single-molecule techniques due to the asynchronous nature of transitions between different molecules of an ensemble. Single-molecule spectroscopy has proven successful in characterizing barrier-crossing dynamics with relatively long average TP times. In the coupled folding and binding of disordered proteins, unusually long-lived non-native contacts lead to TP times on the order of 100 μs [4–6]. The TP time is also relatively long for DNA hairpin folding (20–30 μs). Woodside and coworkers studied the distribution of individual TP times [7] and various other physical properties associated with the TP, including the shape of the TP [8], position dependence of diffusivity [9], and velocity along the TP [10]. However, the TP time is generally short because molecules stay very briefly at the top of the free energy barrier [TP region shown in Fig. 1(a)]. All-atom molecular dynamics (MD) simulations [11] suggest that the folding TP time of small single-domain proteins is on the order of 1 μs, inaccessible by conventional single-molecule spectroscopy. A few TP times close to 10 μs have been measured by slowing down the dynamics using viscogen [12] or for a designed protein [13,14], in which a large number of non-native salt bridges lengthens the TP.

In this Letter, we use a zero-mode waveguide [15] [Fig. 1(b)] that enhances fluorescence emission [16,17] to improve the time resolution by a factor of ∼10 and measure transition path times for protein folding in aqueous solution. We studied eight two-state single-domain proteins (see the Supplemental Material [18], Fig. S1) with different secondary structures (all α, all β, and α/β), sequence lengths, and folding kinetics to investigate how various properties of proteins affect the TP times.

Fluorescence enhancement by nanophotonics—Fluorescence enhancement in single-molecule fluorescence spectroscopy has been achieved using a nanowaveguide [16,17] and DNA origami nanoantennas [40]. We employ the nanowaveguide approach [Fig. 1(b), Supplemental Material [18], Figs. S2 and S3, Appendix A] that does not require immobilization of proteins. Similar to the previous work [17], we observed fluorescence enhancement by a factor of 5–6 [Fig. 1(c)]. At high illumination intensity, many fluorescence bursts exhibit count rates greater than 2 MHz up to 5 MHz [Fig. 1(d)], which increase the time resolution to determine the TP time (see the Supplemental Material [18] for *Single-molecule spectroscopy* and Table S1 for fluorescence burst selection procedure).

*Contact author: chunghoi@niddk.nih.gov

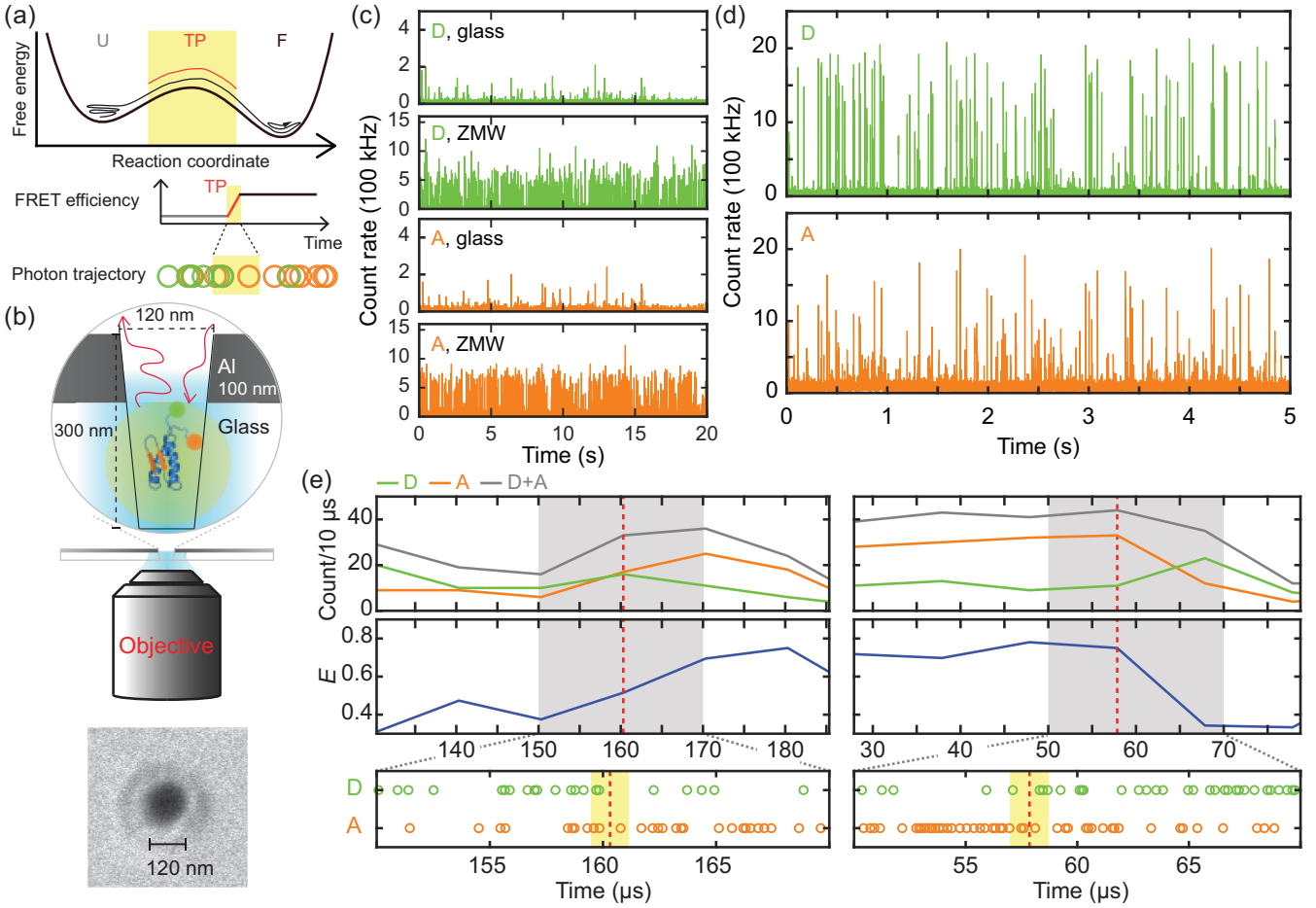


FIG. 1. Single-molecule FRET with zero-mode waveguides to measure folding transition paths. (a) Schematic illustration of the transition paths on a one-dimensional free energy profile (top), idealized (noise-free) FRET efficiency traces (middle), and photon trajectory (bottom). Yellow shaded regions indicate the transition path region. (b) Schematic illustration of the free-diffusion experiment, in which a donor (green circle) and acceptor (orange circle)-labeled protein can diffuse in and out (red trajectory) of a zero-mode waveguide. (Bottom) Scanning electron microscope image of a single zero-mode waveguide. (c) Fluorescence enhancement from the waveguide (0.1 ms bin time). (d) Representative trajectories at high-illumination intensity to measure transition paths (0.1 ms bin time). (e) Representative binned fluorescence trajectories, binned FRET traces, and photon trajectories of a folding transition (left) and an unfolding transition (right). Bin time: 10 μs. Red dashed lines indicate the transition time point determined from the Viterbi algorithm. Yellow shaded regions indicate the transition path region predicted by the maximum likelihood method.

Measurement of the folding TP times of proteins—Figure 2(b) shows the Förster resonance energy transfer (FRET) efficiency histograms for two proteins, villin and Csp*Tm*, which have the fastest and slowest folding kinetics, respectively (see Fig. 5 in Appendix B for the other six proteins). All the measurements were performed near the midpoint of urea denaturation to observe frequent folding and unfolding transitions. The histograms exhibit two peaks (merged due to frequent transitions for villin and well separated for Csp*Tm*) corresponding to the folded and unfolded states (excluding the donor-only species near $E = 0$, gray bars). Photon trajectories of fluorescence bursts were separated into folded and unfolded segments using the Viterbi algorithm [41,42] adapted to photon trajectories [43] with the parameters extracted from the low- and high-illumination intensity data (see the Supplemental Material [18], Tables S2, S3, and Fig. S4). To determine the TP time, we used a kinetic model with an additional virtual intermediate state (S), representing the TP [Fig. 2(a)] [12] (one-intermediate model, see Supplemental Material [18] for *Maximum likelihood analysis and selection of transitions to determine the transition path time*). The statistical significance of the TP time measurement was evaluated by comparing the likelihoods of the one-intermediate model (six-state model, three protein states × acceptor bright or dark states) and the instantaneous transition model (four-state model, two protein states × acceptor bright or dark states) [Fig. 2(a)]. The acceptor dark states had to be included to account for fast acceptor photoblinking events occurring on a similar μs timescale to the TP times. In this initial evaluation, the FRET efficiency of the intermediate state (E_S) was set to the

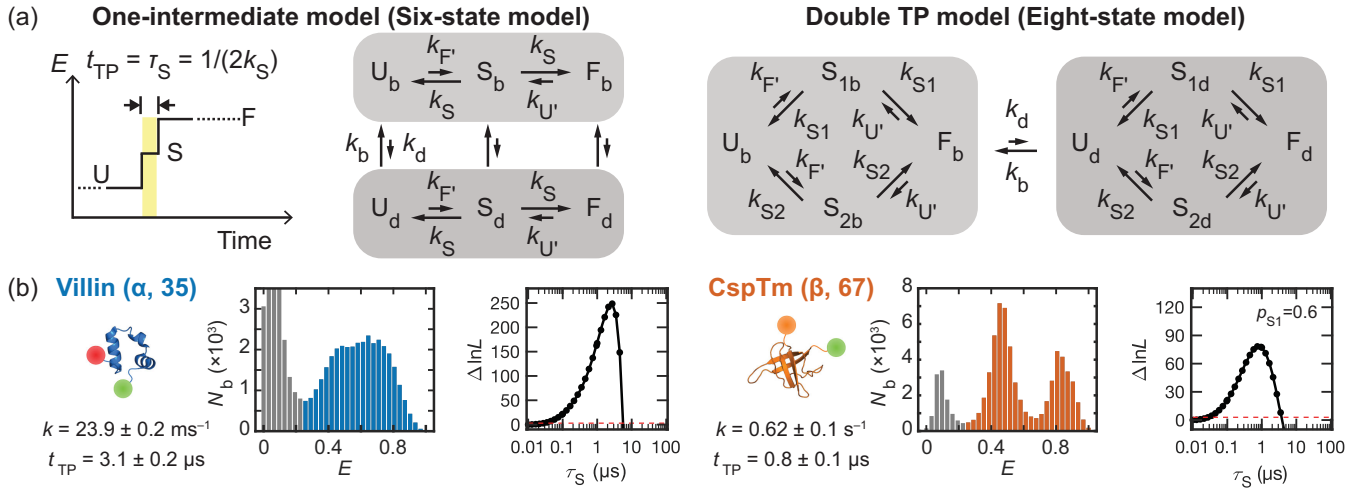


FIG. 2. Maximum likelihood analysis to determine folding TP times. (a) Left: one-intermediate model with acceptor blinking (six-state model). U, unfolded state, S, virtual intermediate state representing the TP highlighted in yellow; F, folded state. The lifetime (τ_S) and FRET efficiency of the intermediate state correspond to the TP time (t_{TP}) and the FRET efficiency, respectively. Subscripted d, acceptor dark state; subscripted b, acceptor bright state. Right: double-TP model (eight-state model). $S1$ and $S2$ represent the true and fake TP, respectively. (b) Left: structures of proteins and their relaxation rates and TP times. Green, orange, and red filled circles indicate Alexa 488, Alexa 594, and Alexa 647. The number of residues are listed in the parenthesis. Middle: the FRET efficiency histograms collected at low-illumination intensity. Colored histogram bars are the FRET species with both donor and acceptor being active, and the gray shaded histogram bars are the donor-only species. Right: log likelihood difference between the six-state model and four-state model (instantaneous transition model, $t_{TP} = 0$) plotted as a function of τ_S, $\Delta \ln L = \ln L(\tau_S) - L(0)$. For protein G and Csp$Tm$, $\Delta \ln L$ was calculated between the eight-state model and six-state model.

average of those of the folded and unfolded states, and all other parameters in the six-state model were fixed to the four-state model parameters (see Supplemental Material [18], Table S4). The log-likelihood differences of the two models, $\Delta \ln L$, are higher than the 95% confidence level [red horizontal dashed line in Fig. 2(b), $\Delta \ln L \approx 3$] across all six fast-folding proteins, assuring the statistical significance of the measurable TP as an additional state [Figs. 2(b) and 5]. In general, the optimal E_S is not exactly the midway value, as shown in the plots with varied E_S values as an example of gpW (see the Supplemental Material [18], Fig. S5). To obtain the accurate TP time (t_{TP}) and E_S, the likelihood function of the six-state model was maximized by varying all parameters (see the Supplemental Material [18], Table S5). The measured TP times of the proteins range from 0.7 to 3.8 μs. For the analysis of the two slow-folding proteins, protein G and CspTm, we used a double-TP model [Fig. 2(a)] to account for the effect of detecting more than one molecule (see Appendix B and the Supplemental Material [18], Figs. S6, S7, and Table S6).

Curvature of a barrier is a key determinant of the TP time—From the measurement of the eight proteins, we explored which properties of proteins affect TP times (t_{TP}). Figure 3(b) shows no dependence of the TP time on the type or content of the secondary structure (all α, all β, and α/β). Figure 3(c) shows an apparent inverse relationship between the protein length and t_{TP}, but our analysis shows that this results mainly from the coincidental correlation between the length and kinetics (i.e., faster kinetics for smaller proteins) as detailed below. All TP times are very short, falling within a narrow range between 700 ns and 4 μs, even though the kinetic folding time (t_F, the average dwell time in the unfolded state before making a folding transition over a barrier, inverse of the folding rate) varies by a factor of 50 000 (80 μs–4 s). It has been predicted that TP time is insensitive to free energy barrier height (a logarithmic dependence on barrier height [12,44,45]), while folding time, t_F, depends exponentially on barrier height [46] as

$$t_F = \frac{1}{k_F} = \frac{2\pi}{\beta D \omega \omega^*} \exp\left(\beta \Delta G_F^*\right), \tag{1}$$

$$t_{TP} \approx \frac{1}{\beta D (\omega^*)^2} \ln\left(2 e^{\gamma} \beta \Delta G_F^*\right). \tag{2}$$

Here, D is the diffusion coefficient at the barrier top, ΔG_F^* is the folding barrier height, $(\omega)^2$ and $(\omega^*)^2$ are the curvatures in the unfolded state and at the barrier top, respectively, $\beta = 1/k_B T$, k_B is the Boltzmann constant, T is absolute temperature, and γ is Euler's constant ($= 0.577\ldots$).

Equations (1) and (2) suggest that TP time is insensitive to folding time. However, Fig. 3(d) clearly shows a trend of an inverse relationship between the t_{TP} and t_F. The TP times of the very fast-folding proteins, WW domain and villin, are longer than those of four fast-folding proteins (protein A, gpW, λ repressor, and ADA2h) and slow-folding proteins

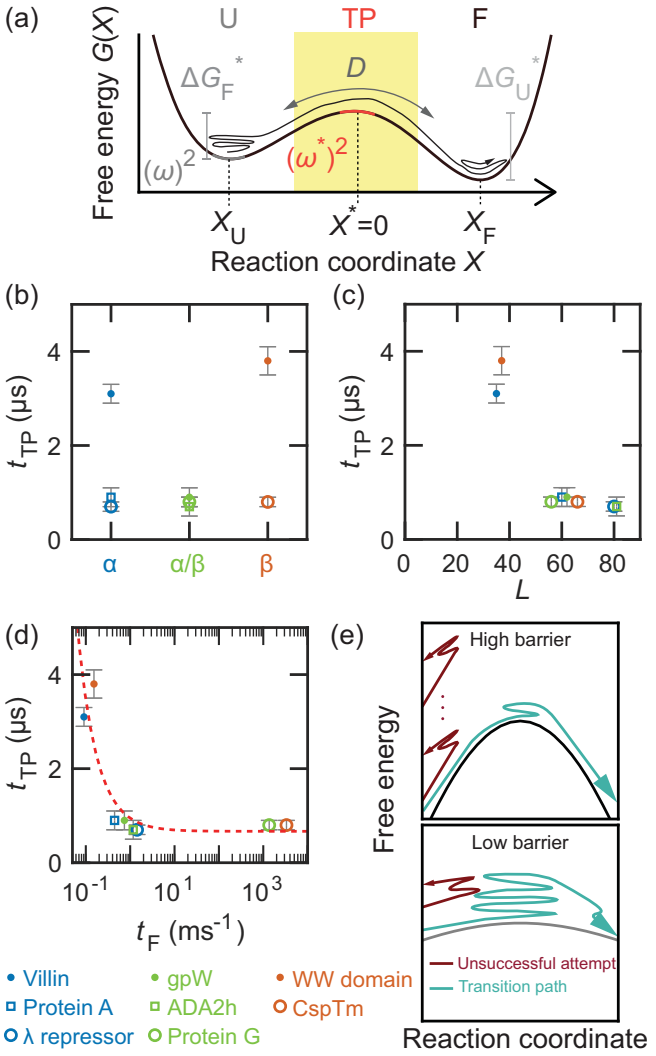


FIG. 3. Dependence of transition path times on properties of proteins. (a) Double-well potential model. D: diffusion coefficient along the one-dimensional reaction coordinate, X. Barrier top is placed at $X^* = 0$. $\Delta G^*_{F,U}$: barrier height of folding and unfolding transitions. $(\omega)^2$: curvature in the unfolded state. $(\omega^*)^2$: curvature at the barrier top. (b) Plot of transition path time t_{TP} against secondary structural archetype (all α, mixed α/β, all β). Gray error bars represent the standard deviation determined from the curvature of the likelihood function at maximum. (c) Relation between t_{TP} and the protein length L (i.e., number of amino acid residues). (d) The relation between the folding time t_F and t_{TP}. Red dashed line indicates an empirical relation between $1/t_F$ and t_{TP} [$t_{TP} = a(1/t_F) + b$]. (e) A simple model to account for the correlation between faster folding kinetics and longer t_{TP}.

(protein G and CspTm). This trend of an inverse relationship may be counterintuitive, but in Eq. (2), two factors affect the TP time more significantly than the barrier height: $(\omega^*)^2$ and D. To evaluate the effect of these factors, we used an empirical model of a one-dimensional (1D) free energy surface [47] with a normalized reaction coordinate (X) similar to the fraction of native contacts (Q) [48–50], a widely used coordinate (see Appendix C and the Supplemental Material [18], Table S7).

Since we use a normalized reaction coordinate, the barrier curvature increases with the barrier height [see Fig. 3(a)]. In other words, the curvature is smaller for fast-folding proteins. If the diffusion coefficient at the barrier is the same across all proteins on the reaction coordinate X (see the next section), the TP times are expected to be longer for fast-folding proteins and shorter for slow-folding proteins [Eq. (2)]. Figure 3(e) provides a physical explanation for why the TP time is shorter for a higher barrier. A molecule makes many more unsuccessful attempts before crossing a high barrier (red trajectories) than a low barrier, resulting in a significant decrease in the folding rate. On the other hand, in the event of barrier crossing (transition path, green trajectories), a molecule spends less time (short diffusion trajectory) at the top of a high barrier because the flat region is narrower than that of a wide, low barrier. To summarize, while the folding time increases exponentially with the barrier height, the TP time decreases monotonically with the barrier height (or curvature). We find that this trend is insensitive to the shape of the barrier (i.e., symmetry) (see the Supplemental Material [18], Table S8). We also find in Fig. 4(a) that the TP time is modulated primarily by the inverse curvature, consistent with the physical picture.

Facilitated diffusion by cooperative native contact formation—The main result of the previous analysis is that the difference in the TP time between proteins is determined by the curvature of the free energy barrier. This analysis relies on the critical assumption that the studied proteins share the same diffusion coefficient on the normalized reaction coordinate X, as suggested by the slope $(1/D)$ in Fig. 4(a). Indeed, the extracted diffusion coefficients in Fig. 4(b) are similar within a factor of 2 (gray shaded area), validating this assumption. However, having a similar diffusion coefficient on the normalized coordinate X is highly counterintuitive. One would expect a constant diffusion coefficient on the number of native contacts (N) coordinate rather than its fraction (Q) because the average stability of a native contact is expected to be similar for all proteins (see Appendix D). The diffusion coefficient along the N coordinate can be obtained by $D_N = D_X \times |dN/dX|^2$, where $|dN/dX|$ is the Jacobian of the coordinate change [51]. By assuming the position-independent (i.e., constant) D_X and D_N along the corresponding coordinates and that N is proportional to X (see Appendix E and the Supplemental Material [18] for *Caveats in the TP time measurements*), $D_N = D_X \times N^2$, which shows a rapid increase with the increasing protein length [Fig. 4(c)] (see Appendix D). In other words, diffusion is facilitated in larger proteins with more native contacts. How is this possible?

Dill and coworkers proposed the "hydrophobic zipper" model [52] to explain protein folding on a reasonable timescale without an exhaustive search of native contacts.

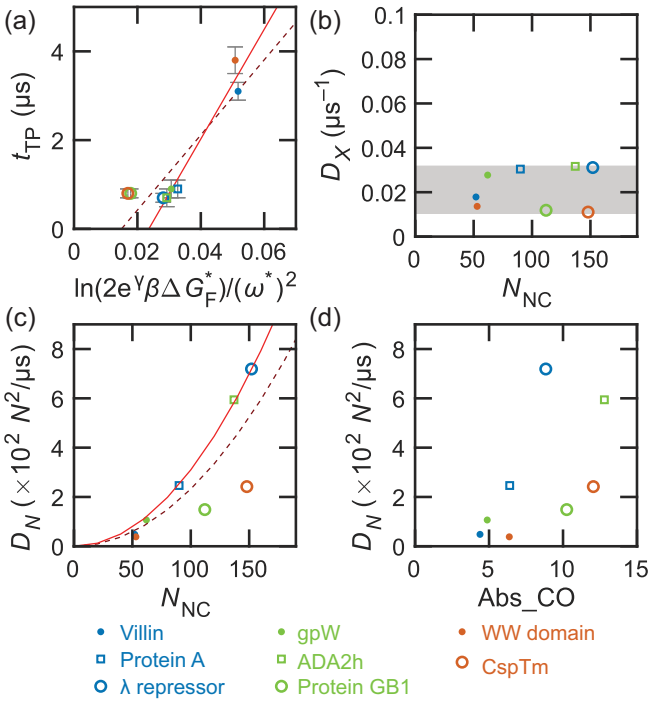


FIG. 4. Relation between diffusion coefficient and native contacts. (a) Relation between t_{TP} and $\ln(2e^{\gamma}\beta\Delta G^{*}_{F})/(\omega^{*})^{2}$ as in the Szabo formula [Eq. (2)]. The slope reports $1/\beta D$. (b) The diffusion coefficient D_X along the normalized coordinate X as a function of the total number of native contacts in the folded structure (N_{NC}). Gray shaded area reflects the range within a factor of 2. (c) Relation between D_N along the number of native contacts coordinate N and N_{NC}. (a),(c) Solid and dashed red curves are linear fits in (a) and quadratic fits in (c) to the six fast-folding proteins and all proteins, including protein G and CspTm, respectively. (d) Relation between D_N and the absolute contact order (Abs_CO).

In this model, a contact formation of two residues separated from each other in a protein sequence (i.e., nonlocal contact formation) facilitates additional nonlocal contacts by bringing more distant residues into proximity, thereby reducing the entropic cost. In coarse-grained simulations by Best and Hummer [51], simultaneous formation and breakage of multiple native contacts were observed in the local barrier crossing events in the transition state and in the folded state, while single or no native contact formation and breakage were observed in the local barrier crossing in the unfolded state. This cooperativity explains the facilitated diffusion on the transition path of larger proteins with more native contacts. Cooperative native contact formation, which was recently observed in the unfolded state in all-atom MD simulations [53], may also play a role in this facilitated diffusion on the TP.

The concept of contact order also provides related interpretations. The contact order (CO) is defined as $\mathrm{CO} = 1/(LN_{NC})\Sigma_{(i,j)}|i-j|$ [see the Supplemental Material [18], Eq. (S26)], where L is the length of the protein (number of amino acid residues) and the summation is performed over the N_{NC} contacts between two residues i and j that are in contact in the folded state [54,55]. It has been shown that CO is highly correlated with $\ln k_F$ among two-state proteins [54]. Absolute contact order [56], $\mathrm{Abs_CO} = L \times \mathrm{CO}$ has also been employed to account for the length dependence of the folding rate [57–59] and to cover a broader range of proteins, including multistate proteins. A higher absolute contact order means that the folded structure has more nonlocal contacts. Figures 4(d), S8, and Table S9 (Supplemental Material [18]) show an increasing trend of D_N with Abs_CO, supporting the idea that facilitated diffusion results from nonlocal long-range contact formation. This is consistent with the hydrophobic zipper model above [52].

The principle of minimal frustration explains how natural proteins can fold in a biologically relevant timescale by optimizing the energy landscape to avoid local traps formed by non-native interactions [60]. The folding TP times for molecules, for which energy landscape optimization may be inefficient, such as nucleic acids and designed proteins, and non-native protein dimers are very long, ranging from tens of microseconds to a millisecond (see Appendix F). The much shorter TP times and the increasing diffusion coefficient with the length of natural proteins suggest that facilitated diffusion is another remarkable result of reduced landscape roughness through evolutionary optimization.

Acknowledgments—We thank W. A. Eaton, A. Szabo, R. B. Best, and I. V. Gopich for numerous helpful discussions and comments; A. Aniana for technical assistance associated with the preparation of recombinant proteins; and M. Treviño for collecting preliminary data to determine the urea midpoint of λ repressor, protein A, and WW domain. Fabrication of zero-mode waveguides was supported by the NIH Multi-Institute Cryo-EM Facility (MICEF). This research was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) within the National Institutes of Health (NIH). The contributions of the NIH authors are considered Works of the United States Government.

C.-J. F. and H. S. C. designed research. C.-J. F., H. S. C., and J. M. L. wrote the manuscript. C.-J. F. performed research and data analysis. C.-J. F. and U. B. fabricated nanowaveguides. J. M. L. expressed and purified dye-labeled proteins.

The findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

Data availability—The experimental data and analysis that support the findings of this article are openly available [61].

The source codes and compiled libraries and analysis routines using MATLAB for the transition path analysis, instructions for optimization using a multi-thread CPU calculation [62], and example data are available in the GitHub repository [63]. Materials used in this study are available from H. S. C. under a material transfer agreement with the National Institutes of Health.

[1] H. S. Chung and W. A. Eaton, Protein folding transition path times from single molecule FRET, Curr. Opin. Struct. Biol. **48**, 30 (2018).

[2] H. S. Chung, Transition path times measured by single-molecule spectroscopy, J. Mol. Biol. **430**, 409 (2018).

[3] N. Q. Hoffer and M. T. Woodside, Probing microscopic conformational dynamics in folding reactions by measuring transition paths, Curr. Opin. Chem. Biol. **53**, 68 (2019).

[4] J.-Y. Kim, F. Meng, J. Yoo, and H. S. Chung, Diffusion-limited association of disordered protein by non-native electrostatic interactions, Nat. Commun. **9**, 4707 (2018).

[5] F. Sturzenegger, F. Zosel, E. D. Holmstrom, K. J. Buholzer, D. E. Makarov, D. Nettels, and B. Schuler, Transition path times of coupled folding and binding reveal the formation of an encounter complex, Nat. Commun. **9**, 4708 (2018).

[6] J.-Y. Kim and H. S. Chung, Disordered proteins follow diverse transition paths as they fold and bind to a partner, Science **368**, 1253 (2020).

[7] K. Neupane, D. A. N. Foster, D. R. Dee, H. Yu, F. Wang, and M. T. Woodside, Direct observation of transition paths during the folding of proteins and nucleic acids, Science **352**, 239 (2016).

[8] N. Q. Hoffer, K. Neupane, A. G. T. Pyo, and M. T. Woodside, Measuring the average shape of transition paths during the folding of a single biological molecule, Proc. Natl. Acad. Sci. U.S.A. **116**, 8125 (2019).

[9] D. A. N. Foster, R. Petrosyan, A. G. T. Pyo, A. Hoffmann, F. Wang, and M. T. Woodside, Probing position-dependent diffusion in folding reactions using single-molecule force spectroscopy, Biophys. J. **114**, 1657 (2018).

[10] K. Neupane, N. Q. Hoffer, and M. T. Woodside, Measuring the local velocity along transition paths during the folding of single biological molecules, Phys. Rev. Lett. **121**, 018102 (2018).

[11] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, How fast-folding proteins fold, Science **334**, 517 (2011).

[12] H. S. Chung, K. McHale, J. M. Louis, and W. A. Eaton, Single-molecule fluorescence experiments determine protein folding transition path times, Science **335**, 981 (2012).

[13] H. S. Chung and W. A. Eaton, Single-molecule fluorescence probes dynamics of barrier crossing, Nature (London) **502**, 685 (2013).

[14] H. S. Chung, S. Piana-Agostinetti, D. E. Shaw, and W. A. Eaton, Structural origin of slow diffusion in protein folding, Science **349**, 1504 (2015).

[15] M. J. Levene, J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead, and W. W. Webb, Zero-mode waveguides for single-molecule analysis at high concentrations, Science **299**, 682 (2003).

[16] M. Baibakov, S. Patra, J.-B. Claude, A. Moreau, J. Lumeau, and J. Wenger, Extending single-molecule Förster resonance energy transfer (FRET) range beyond 10 nanometers in zero-mode waveguides, ACS Nano **13**, 8469 (2019).

[17] M. F. Nüesch, M. T. Ivanović, J.-B. Claude, D. Nettels, R. B. Best, J. Wenger, and B. Schuler, Single-molecule detection of ultrafast biomolecular dynamics with nanophotonics, J. Am. Chem. Soc. **144**, 52 (2022).

[18] See Supplemental Material at http://link.aps.org/supplemental/10.1103/2q9m-4scc for experimental methods, data analysis, simulation of photon trajectories, additional discussion, 11 Supplementary Figures, and 11 Supplementary Tables, which includes Refs. [19–39].

[19] B. Schuler, E. A. Lipman, and W. A. Eaton, Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy, Nature (London) **419**, 743 (2002).

[20] H. S. Chung, J. M. Louis, and I. V. Gopich, Analysis of fluorescence lifetime and energy transfer efficiency in single-molecule photon trajectories of fast-folding proteins, J. Phys. Chem. B **120**, 680 (2016).

[21] H. S. Chung, T. Cellmer, J. M. Louis, and W. A. Eaton, Measuring ultrafast protein folding rates from photon-by-photon analysis of single molecule fluorescence trajectories, Chem. Phys. **422**, 229 (2013).

[22] H. Gouda, H. Torigoe, A. Saito, M. Sato, Y. Arata, and I. Shimada, Three-dimensional solution structure of the B domain of staphylococcal protein A: Comparisons of the solution and crystal structures, Biochemistry **31**, 9665 (1992).

[23] L. J. Beamer and C. O. Pabo, Refined 1.8 Å crystal structure of the λ repressor-operator complex, J. Mol. Biol. **227**, 177 (1992).

[24] M. A. Jiménez, V. Villegas, J. Santoro, L. Serrano, J. Vendrell, F. X. Avilés, and M. Rico, NMR solution structure of the activation domain of human procarboxypeptidase A2, Protein Sci. **12**, 296 (2003).

[25] S. Patra, M. Baibakov, J.-B. Claude, and J. Wenger, Surface passivation of zero-mode waveguide nanostructures: Benchmarking protocols and fluorescent labels, Sci. Rep. **10**, 5235 (2020).

[26] S. Yang, N. Klughammer, A. Barth, M. E. Tanenbaum, and C. Dekker, Zero-mode waveguide nanowells for single-molecule detection in living cells, ACS Nano **17**, 20179 (2023).

[27] S. Tiwari, P. Roy, J. Claude, and J. Wenger, Achieving high temporal resolution in single-molecule fluorescence techniques using plasmonic nanoantennas, Adv. Opt. Mater. **11**, 2300168 (2023).

[28] J. Vogelsang, R. Kasper, C. Steinhauer, B. Person, M. Heilemann, M. Sauer, and P. Tinnefeld, A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes, Angew. Chem., Int. Ed. **47**, 5465 (2008).

[29] T. M. P. Hartwich, K. K. Hin Chung, L. Schroeder, J. Bewersdorf, C. Soeller, and D. Baddeley, A stable, high refractive index, switching buffer for super-resolution imaging, bioRxiv, 465492 (2018), 10.1101/465492.

[30] C. Eggeling, S. Berger, L. Brand, J. R. Fries, J. Schaffer, A. Volkmer, and C. A. M. Seidel, Data registration and selective single-molecule analysis using multi-parameter fluorescence detection, J. Biotechnol. **86**, 163 (2001).

[31] I. V. Gopich, J.-Y. Kim, and H. S. Chung, Analysis of photon trajectories from diffusing single molecules, J. Chem. Phys. **159**, 024119 (2023).
[32] G. Hummer, From transition paths to transition states and rate coefficients, J. Chem. Phys. **120**, 516 (2004).
[33] F. Liu, M. Nakaema, and M. Gruebele, The transition state transit time of WW domain folding is controlled by energy landscape roughness, J. Chem. Phys. **131**, 195101 (2009).
[34] W. Y. Yang and M. Gruebele, Folding at the speed limit, Nature (London) **423**, 193 (2003).
[35] F. Liu, D. Du, A. A. Fuller, J. E. Davoren, P. Wipf, J. W. Kelly, and M. Gruebele, An experimental survey of the transition between two-state and downhill protein folding scenarios, Proc. Natl. Acad. Sci. U.S.A. **105**, 2369 (2008).
[36] H. S. Chan, Z. Zhang, S. Wallin, and Z. Liu, Cooperativity, local-nonlocal coupling, and nonnative interactions: Principles of protein folding from coarse-grained models, Annu. Rev. Phys. Chem. **62**, 301 (2011).
[37] R. B. Best, N.-V. Buchete, and G. Hummer, Are current molecular dynamics force fields too helical?, Biophys. J. **95**, L07 (2008).
[38] S. Piana, K. Lindorff-Larsen, and D. E. Shaw, How robust are protein folding simulations with respect to force field parameterization?, Biophys. J. **100**, L47 (2011).
[39] S. Nauli, B. Kuhlman, and D. Baker, Computer-based redesign of a protein folding pathway, Nat. Struct. Biol. **8**, 602 (2001).
[40] L. Grabenhorst, F. Sturzenegger, M. Hasler, B. Schuler, and P. Tinnefeld, Single-molecule FRET at 10 MHz count rates, J. Am. Chem. Soc. **146**, 3539 (2024).
[41] A. Viterbi, Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, IEEE Trans. Inf. Theory **13**, 260 (1967).
[42] L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE **77**, 257 (1989).
[43] H. S. Chung, I. V. Gopich, K. McHale, T. Cellmer, J. M. Louis, and W. A. Eaton, Extracting rate coefficients from single-molecule photon trajectories and FRET efficiency histograms for a fast-folding protein, J. Phys. Chem. A **115**, 3642 (2011).
[44] H. S. Chung, J. M. Louis, and W. A. Eaton, Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories, Proc. Natl. Acad. Sci. U.S.A. **106**, 11837 (2009).
[45] H. S. Chung and I. V Gopich, Fast single-molecule FRET spectroscopy: Theory and experiment, Phys. Chem. Chem. Phys. **16**, 18644 (2014).
[46] H. A. Kramers, Brownian motion in a field of force and the diffusion model of chemical reactions, Physica (Amsterdam) **7**, 284 (1940).
[47] N. Mothi and V. Munõz, Protein folding dynamics as diffusion on a free energy surface: Rate equation terms, transition paths, and analysis of single-molecule photon trajectories, J. Phys. Chem. B **125**, 12413 (2021).
[48] B. C. Gin, J. P. Garrahan, and P. L. Geissler, The limited role of nonnative contacts in the folding pathways of a lattice protein, J. Mol. Biol. **392**, 1303 (2009).
[49] R. B. Best, G. Hummer, and W. A. Eaton, Native contacts determine protein folding mechanisms in atomistic simulations, Proc. Natl. Acad. Sci. U.S.A. **110**, 17874 (2013).
[50] W. Zheng and R. B. Best, Reduction of all-atom protein folding dynamics to one-dimensional diffusion, J. Phys. Chem. B **119**, 15247 (2015).
[51] R. B. Best and G. Hummer, Coordinate-dependent diffusion in protein folding, Proc. Natl. Acad. Sci. U.S.A. **107**, 1088 (2010).
[52] K. A. Dill, K. M. Fiebig, and H. S. Chan, Cooperativity in protein-folding kinetics, Proc. Natl. Acad. Sci. U.S.A. **90**, 1942 (1993).
[53] D. Wang, L. B. Frechette, and R. B. Best, On the role of native contact cooperativity in protein folding, Proc. Natl. Acad. Sci. U.S.A. **121**, e2319249121 (2024).
[54] K. W. Plaxco, K. T. Simons, and D. Baker, Contact order, transition state placement and the refolding rates of single domain proteins, J. Mol. Biol. **277**, 985 (1998).
[55] A. R. Dinner and M. Karplus, The roles of stability and contact order in determining protein folding rates, Nat. Struct. Biol. **8**, 21 (2001).
[56] D. N. Ivankov, S. O. Garbuzynskiy, E. Alm, K. W. Plaxco, D. Baker, and A. V. Finkelstein, Contact order revisited: Influence of protein size on the folding rate, Protein Sci. **12**, 2057 (2003).
[57] P. G. Wolynes, Folding funnels and energy landscapes of larger proteins within the capillarity approximation, Proc. Natl. Acad. Sci. U.S.A. **94**, 6170 (1997).
[58] M. S. Li, D. K. Klimov, and D. Thirumalai, Thermal denaturation and folding rates of single domain proteins: Size matters, Polymer **45**, 573 (2004).
[59] A. N. Naganathan and V. Muñoz, Scaling of folding times with protein size, J. Am. Chem. Soc. **127**, 480 (2005).
[60] J. D. Bryngelson and P. G. Wolynes, Spin glasses and the statistical mechanics of protein folding, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).
[61] C.-J. Feng, U. Baxa, J. M. Louis, and H. S. Chung, Single-molecule FRET protein transition path measurement data, Zenodo, 10.5281/zenodo.18354860.
[62] J. Yoo, J.-Y. Kim, J. M. Louis, I. V. Gopich, and H. S. Chung, Fast three-color single-molecule FRET using statistical inference, Nat. Commun. **11**, 3336 (2020).
[63] https://github.com/hoisunglab/FRET_TransitionPath.
[64] J. Kubelka, J. Hofrichter, and W. A. Eaton, The protein folding "speed limit", Curr. Opin. Struct. Biol. **14**, 76 (2004).
[65] G. H. Taumoefolau and R. B. Best, Estimating transition path times and shapes from single-molecule photon trajectories: A simulation analysis, J. Chem. Phys. **154**, 115101 (2021).
[66] I. V Gopich and A. Szabo, Decoding the pattern of photon colors in single-molecule FRET, J. Phys. Chem. B **113**, 10965 (2009).
[67] K. Neupane, F. Wang, and M. T. Woodside, Direct measurement of sequence-dependent transition path times and conformational diffusion in DNA duplex formation, Proc. Natl. Acad. Sci. U.S.A. **114**, 1329 (2017).
[68] K. Neupane, D. B. Ritchie, H. Yu, Daniel A. N. Foster, F. Wang, and M. T. Woodside, Transition path times for nucleic acid folding determined from energy-landscape

analysis of single-molecule trajectories, Phys. Rev. Lett. **109**, 068102 (2012).

[69] K. Truex, H. S. Chung, J. M. Louis, and W. A. Eaton, Testing landscape theory for biomolecular processes with single molecule fluorescence spectroscopy, Phys. Rev. Lett. **115**, 018101 (2015).

[70] H. Yu, A. N. Gupta, X. Liu, K. Neupane, A. M. Brigley, I. Sosova, and M. T. Woodside, Energy landscape analysis of native folding of the prion protein yields the diffusion constant, transition path time, and rates, Proc. Natl. Acad. Sci. U.S.A. **109**, 14452 (2012).

End Matter

Appendix A: Nanowaveguides experiment—A conventional confocal free-diffusion experiment is straightforward because it does not rely on the exact position of the laser focus. In contrast, a waveguide experiment requires the focus to stay within 20 nm to keep high fluorescence emission. To this end, the system was automatically refocused by a three-dimensional scan of the surface periodically every minute (see the Supplemental Material [18] for *Automatic correction of focus drift during free-diffusion experiment in waveguides* and Fig. S3). In our experiments, several thousand analyzable fluorescence bursts could be collected in 1–2 h for six relatively fast-folding proteins (relaxation rate 1.6–26 ms^{-1}). However, the folding rates of two slow folders, protein G and Csp*Tm* are 1.5 and 0.6 s^{-1}, respectively, which requires 1000 times longer data collection time to obtain a comparable number of transitions. Overall, it took 51 days (data collection time of 34 days, total burst duration 2.5 h) to collect 19 605 transitions for protein G and 37 days (data collection time of 29 days, total burst duration 2.8 h) to collect 20 562 transitions for Csp*Tm*. Therefore, the implementation of autofocusing was particularly critical for the measurement of these two slow-folding proteins.

Appendix B: Analysis of proteins with slow folding kinetics—For protein G and Csp*Tm*, the $\Delta \ln L$ plots using the six-state model show unexpectedly long TP times of 7–10 μs [see the Supplemental Material [18], Figs. S6(a) and S7(a)], which is almost 1 order of magnitude longer than other fast-folding proteins. We suspect that this long TP time results from the detection of multiple molecules in the waveguide. Given the experimental concentrations of 35 nM, the average number of molecules occupying the waveguide is 0.05, which ensures that most of the transitions for the fast-folding proteins result from true folding or unfolding transitions. In contrast, the observed transitions for protein G (19 605) and Csp*Tm* (20 562) far exceed those expected from the total burst durations and relaxation rates (13 500 and 6250, respectively). This result indicates that the analysis is significantly impacted by fake transitions resulting from the diffusion of two molecules in and out of the waveguide. For example, a long transition path time can be observed if an unfolded molecule diffuses into the waveguide when a folded molecule is already inside, and then the folded molecule diffuses out. Other possible scenarios of detecting fake transitions are listed in Table S6 (Supplemental Material [18]). In most cases, these fake transitions result in a long transition path time. To determine the true TP times, we employed a double-TP model [eight-state model, Fig. 2(a) and Eq. (S8), Supplemental Material [18]]. The two virtual intermediate states, S1 and S2, account for the true TPs and fake TPs exhibiting long TP times by multiple molecules, respectively. Figure S6(b) (see the Supplemental Material [18]) shows a series of $\Delta \ln L$ plots with a varied relative fraction of the two TPs of protein G. The parameters for the long TP were fixed to the values obtained from the six-state model. The highest likelihood value is found at $t_{TP} \sim 1$ μs when the short TP population $p_{S1} = 0.7$ [Fig. 2 and Supplemental Material [18], Fig. S6(b)], suggesting the presence of a very short (true) TP time. Similar behaviors are observed in Csp*Tm* (see the Supplemental Material [18], Fig. S7). To determine the parameters accurately, we fitted the data with free parameters of the double-TP model and obtained the TP times of these proteins from the value of the fast TP (S1) (see Supplemental Material [18], Table S5). These times are 0.8 μs for both protein G and Csp*Tm* [Figs. 2(b) and 5]. The parameters of the double-TP model for the six fast-folding proteins converge to those of the single-TP model (six-state), indicating that the effect of fake transitions is negligible in the data analysis of these proteins.

Appendix C: One-dimensional free energy surface—Evaluating the effect of the curvature of the barrier, $(\omega^*)^2$, and the diffusion coefficient, D, on TP times requires defining the folding reaction coordinate. Since it has been shown that protein folding can be described very well by diffusion on a one-dimensional (1D) free energy surface with the fraction of native contacts (Q) as the reaction coordinate [48–50], we used an empirical model of the free energy surface [47] with a similarly normalized reaction coordinate (X) [see Fig. 3(a) and the Supplemental Material [18] for *Estimating barrier height and diffusion coefficient using a double-well potential model*]. The separation between the minima of the folded and unfolded wells is 1 in this coordinate. This model allows an asymmetric barrier, and the FRET efficiency values of the folded, unfolded, and the TP can

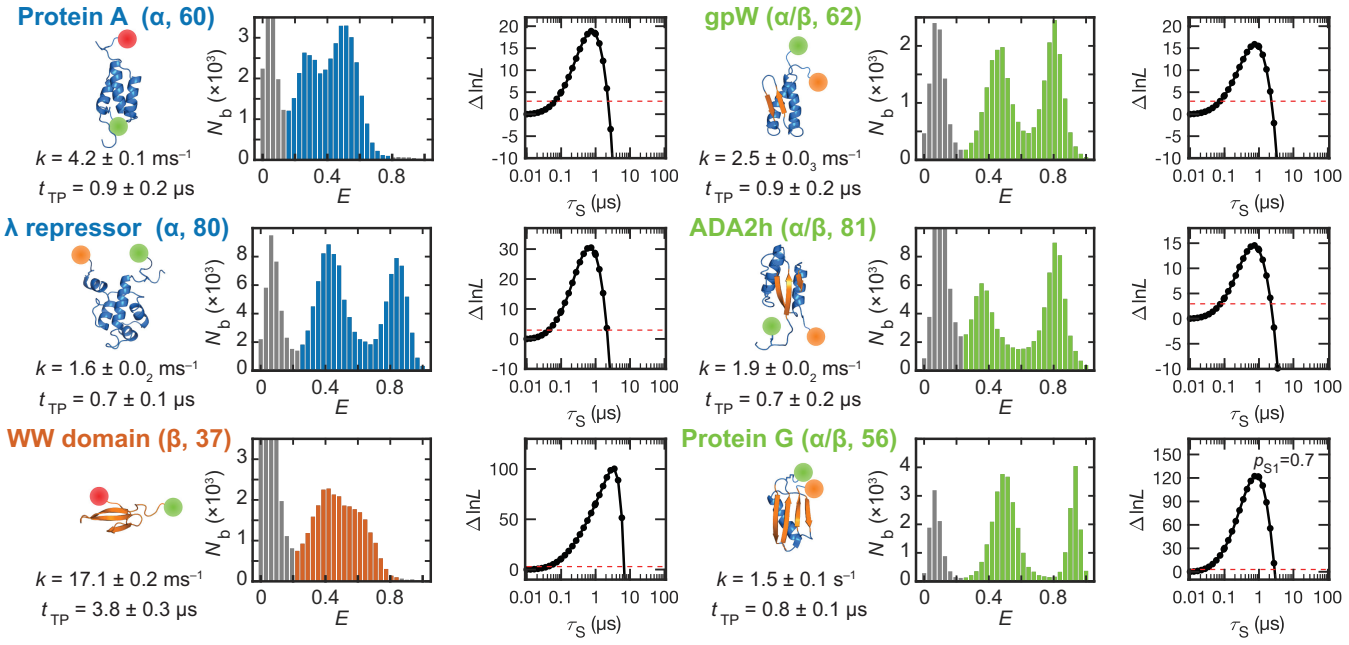


FIG. 5. Maximum likelihood analysis to determine folding TP times. Left: structures of proteins and their relaxation rates and TP times. The secondary structures and number of residues are listed in the parenthesis. Middle: the FRET efficiency histograms collected at low illumination intensity. Right: log-likelihood difference, $\Delta \ln L = \ln L(\tau_S) - L(0)$, plotted as a function of τ_S.

be used to estimate the asymmetry (i.e., position of the free energy barrier). From the measured k_F and t_{TP} it is possible to extract the barrier height, the diffusion coefficient, and barrier curvature that depends on barrier height and separation of the wells [Fig. 3(a); see the Supplemental Material [18], Table S7, *Estimating barrier height and diffusion coefficient using a double-well potential model*]. Note that these properties cannot be obtained by measuring the folding kinetics alone because the preexponential factor in Eq. (1) is unknown.

Appendix D: Length dependence of diffusion coefficient—The insensitivity of D_X to protein length observed in our experiment is inconsistent with the analysis results of MD simulations [50] that showed D actually decreases with the length ($D_Q \propto L^{-2.1}$). However, excluding the data of a short ten-residue peptide weakens the length dependence. Some of the discrepancy may be attributed to factors including the high simulation temperature and imperfect force field (see the Supplemental Material [18] for *Comparison of the length dependence of diffusivity in simulation and experiment* for further discussion). Kubelka *et al.* [64] estimated that the folding speed limit, which corresponds to the preexponential factor of Kramers rate equation [Eq. (1)] and relates to the TP time, is proportional to the length L. However, this estimation assumed that the diffusion coefficient at the barrier top is the same as that in the unfolded state, suggesting that the dependence of D on L may differ between the unfolded state and at the barrier top. Position-dependent D has also been reported in the analysis of MD simulations [50,51]. Our analysis assumes a constant D over the free energy barrier.

Appendix E: Accuracy of TP time measurement—The findings in this Letter rely on an accurate determination of the TP times. Compared to the folding kinetics, the TP time is extremely difficult to measure due to its short duration and low population. For instance, based on the TP time and folding and unfolding rates, the fraction of molecules in the TP is estimated to be 4% for the fastest folder, villin, and 0.0001% for the slowest folder, Csp*Tm*. The accuracy of the TP time measured in this Letter depends on the nature of the measured quantity (FRET efficiency) and the validity of the TP model (i.e., one-intermediate model) and statistical analysis, all of which have caveats. These include underestimation of the TP time by a factor of 2 due to nonoptimal distance-based reaction coordinate X rather than Q or N [65], (see the Supplemental Material [18] for *Caveats in the TP time measurements*, Fig. S9 and Table S10), associated effect from barrier asymmetry (see the Supplemental Material [18], Fig. S10) and coordinate conversion from X to N, potential systematic underestimation of TP times by a factor of 1.7 (see the Supplemental Material [18] for *Recoloring simulation of photon trajectories* [66], Fig. S11 and Table S11), and overestimation of the TP times of protein G and Csp*Tm* due to incomplete subtraction of the slow TP component (see the Supplemental Material [18] for *Caveats in the TP time measurements*).

Appendix F: Comparison of folding TP times of biomolecules—Despite several caveats in the experimental determination of the TP times and potential underestimation, our data show a robust dependence of the extracted diffusion coefficient on the number of native contacts or protein length. These results imply that diffusion in large proteins may be facilitated as a result of evolutionary optimization of the folding energy landscape. Compared to natural proteins, efficient optimization may not be possible in a designed protein, exhibiting long TP times [13,14], and in nucleic acids because there are only four different nucleotides and two types of base pairing. Single-molecule force spectroscopy of 20-bp DNA hairpin showed the sensitivity of the TP times and diffusion to the sequence composition (content of guanine-cytosine pairs) [67], suggesting inefficient optimization. Their TP times of 20–30 μs, an order of magnitude longer than those of the proteins studied in this Letter, also indicate less optimization. In addition, the TP time of DNA hairpins increases with the length of base pairs [68,69], different from the trend observed in protein folding in this Letter. Finally, the transition path time of the non-native dimer formation of the prion protein was observed to be close to a millisecond [7], much longer than that of the monomer folding, 2 μs, obtained from the energy landscape deconvolution [70].