THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Protein Folding Dynamics as Diffusion on a Free Energy Surface: Rate Equation Terms, Transition Paths, and Analysis of Single-Molecule Photon Trajectories

Published as part of The Journal of Physical Chemistry virtual special issue "Dave Thirumalai Festschrift".

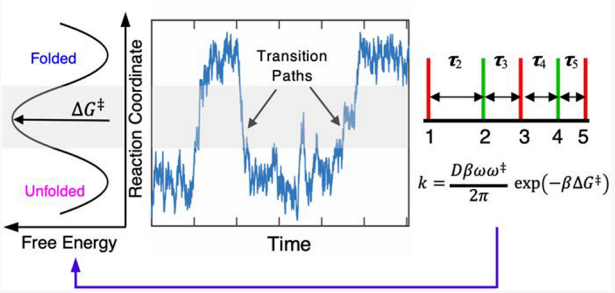
Nivin Mothi and Victor Muñoz*

Cite This: *J. Phys. Chem. B* 2021, 125, 12413−12425

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The rates of protein (un)folding are often described as diffusion on the projection of a hyperdimensional energy landscape onto a few (ideally one) order parameters. Testing such an approximation by experiment requires resolving the reactive transition paths of individual molecules, which is now becoming feasible with advanced single-molecule spectroscopic techniques. This has also sparked the interest of theorists in better understanding reactive transition paths. Here we focus on these issues aiming to establish (i) practical guidelines for the mechanistic interpretation of transition path times (TPT) and (ii) methods to extract the free energy surface and protein dynamics from the maximum likelihood analysis of photon trajectories (MLA-PT). We represent the (un)folding rates as diffusion on a 1D free energy surface with the FRET efficiency as a reaction coordinate proxy. We then perform diffusive kinetic simulations on surfaces with two minima and a barrier, but with different shapes (curvatures, barrier height, and symmetry), coupled to stochastic simulations of photon emissions that reproduce current SM-FRET experiments. From the analysis of transition paths, we find that the TPT is inversely proportional to the barrier height (difference in free energy between minimum and barrier top) for any given surface shape, and that dividing the TPT into climb and descent segments provides key information about the barrier's symmetry. We also find that the original MLA-PT procedure used to determine the TPT from experiments underestimates its value, particularly for the cases with smaller barriers (e.g., fast folders), and we suggest a simple strategy to correct for this bias. Importantly, we also demonstrate that photon trajectories contain enough information to extract the 1D free energy surface's shape and dynamics (if TPT is >4−5-fold longer than the interphoton time) using the MLA-PT directly implemented with a diffusive free energy surface model. When dealing with real (unknown) experimental data, the comparison between the likelihoods of the free energy surface and discrete kinetic three-state models can be used to evaluate the statistical significance of the estimated free energy surface.

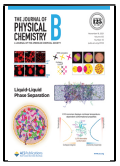


INTRODUCTION

Protein folding rates are determined by the stochastic search for the native state on a hyperdimensional phase space defined by the coordinates of all the protein atoms and surrounding solvent molecules.[1] The theoretical expectation is hence that individual protein molecules follow myriads of microscopic paths, resulting in broad distributions of mechanisms.[2,3] Multiple parallel pathways have been observed in coarse-grained[4−6] and atomistic simulations.[7−10] However, such expected microscopic heterogeneity is inaccessible to conventional bulk kinetic experiments, which, regardless of temporal resolution, only report on the collective relaxation rate. Molecular dynamics simulations in explicit solvent, which provide high-resolution 3D structures in picosecond time intervals from single trajectories,[11] and now reach milliseconds,[12−14] may ultimately provide such detailed information. However, the imperfect/approximate nature of the physical force fields employed in these simulations make it essential to benchmark their results with experiments. In this regard, theory also proposes that the rate can be described as diffusion on the projection of the energy landscape onto a few (even one) order parameter(s),[15] thus giving rise to a Kramers-like rate expression:[16]

$$k = \left(\frac{D\beta\omega\omega^{\ddagger}}{2\pi}\right) \exp(-\beta\Delta G^{\ddagger}) \tag{1}$$

Received: June 18, 2021
Revised: October 12, 2021
Published: November 4, 2021



where D is the intramolecular diffusion coefficient, ω^2 is the curvature of the well (e.g., folded and unfolded), $-(\omega^{\ddagger})^2$ is the inverted curvature at the barrier top, and $\Delta G^{\ddagger}$ is the free energy barrier on the projected free energy surface (FES). All these terms become empirical (not derivable from first-principles) and contain a great deal of mechanistic information, which makes them dependent on the projection,[17,18] the structure of the protein, and its sequence.[19] This approach also raises other important issues, such as the emergence of free energy barriers caused by entropic bottlenecks,[20] whether the intramolecular diffusion coefficient (D) is invariant along the reaction coordinate,[21,22] or which and how many order parameters are needed to conform an effective reaction coordinate.[23]

It is important to note that the terms of even a simplified rate expression like eq 1 remain inaccessible to bulk kinetic experiments, which accordingly have not been able to directly test theoretical predictions or to extract much mechanistic information from the rates.[24] Fast folding experiments[25,26] provided a first exposure to these issues by (i) measuring the sub-microsecond time scales of elementary protein conformational motions,[27,28] (ii) detecting an additional, superfast kinetic ("molecular") phase on sub-millisecond folding proteins that reflects the relaxation from the barrier top,[29] and (iii) identifying and characterizing the time scales of barrierless (downhill) folding.[30] Those results led to estimates of the overall pre-exponential in eq 1, or folding speed limit,[31] which not only enabled the thermodynamic analysis of experimental rates and confirmed the entropic origin of folding barriers[32] but also raised new questions about the source of the rate's temperature dependence.[33] Comparative analysis of the "molecular" phase has recently shown that the overall pre-exponential term can change by at least 10-fold on fast folders of the same size and fold.[34]

Advances from fast-folding techniques notwithstanding, the only way to experimentally dissect (un)folding rates is by detecting individual molecules with sufficient resolution to resolve the reactive transition paths.[35] Transition paths are the segments of a stochastic trajectory during which the protein crosses the free energy barrier. It has been shown that the shape of transition paths and their distribution can provide sufficient information to dissect the rate expression[23,36,37] and hence to infer the underlying mechanisms.[38] There currently are two options to access transition paths by experiment: single-molecule Förster resonance energy transfer (SM-FRET)[39] and force spectroscopy (FS).[40] Recent FS experiments and simulations have provided strong evidence,[41,42] and direct observation,[43] of multiple mechanical unfolding paths in otherwise "two-state" folding proteins. Mechanical (un)folding transition paths have recently been resolved using optical tweezers[44,45] and atomic force microscopy.[46] However, in mechanical unfolding experiments the molecule is tethered to a pulling device (beads or cantilever) that introduces strong dynamic effects[47] for which one must account.[48] In parallel, the implementation of MLA-PT methods[49] has led to major advances in the effective time resolution of SM-FRET. For instance, MLA-PT combined with elementary (two- and three-state) kinetic models has been used to determine fast (un)folding rate coefficients (k)[50] and microsecond mean transition path times (TPT) on several small proteins.[35,51,52] TPTs alone do not contain enough information to extract the FES or D,[37,53] but the ratio between TPT and rate coefficient has been used to estimate the barrier and overall pre-exponential term.[35] Such analysis approximates the TPT to an analytical expression for a harmonic barrier $>2k_BT$ developed by Szabo:[38,54]

$$\langle \tau_{TP} \rangle \approx \ln[2e^{\gamma}\beta\Delta G^{\ddagger}]/D\beta(\omega^{\ddagger})^2 \quad (2)$$

where γ is Euler's constant (0.577...) and $(\omega^{\ddagger})^2$ is the curvature of the barrier. It also involves the assumption of equal curvature for the wells and barrier, so they can be eliminated from the ratio. The same general approach has been recently extended to folding upon binding transitions of IDPs in which TPTs seem to be somewhat slower.[55,56]

This methodology for the analysis of SM-FRET experiments raises some interesting questions, such as: How accurate are the TPTs so estimated? How do they depend on the various terms of the rate expression? Ultimately, is the approach potentially extensible to directly determine the shape of transition paths and hence probe the underlying folding mechanisms? In previous work, we explored the performance of MLA-PT in extracting a known 1D free energy surface of fixed shape and diffusion coefficient from simulated photon trajectories coupled to intramolecular diffusion on that surface.[57] That study showed that the method could extract the barrier height and diffusion coefficient when the shape of the free energy surface is predefined.[57] Very recently, Taumoefolau and Best have used coarse-grained folding simulations to assess the suitability of the FRET efficiency (E) as reaction coordinate and of the two-/three-state MLA-PT procedure to estimate the TPT.[58] In that case, the authors compared the results from the simulated photon trajectories with the TPT obtained by projecting the molecular trajectories as a function of the fraction of native contacts (Q), an order parameter that is generally accepted as a suitable folding reaction coordinate.[10,15,21]

Here we investigate the relationships between the shape of the free energy surface, the reactive transition paths, and the performance of various MLA-PT methods. We start by assuming that the (un)folding process can be described as diffusion on a 1D FES and that E is a reasonable reaction coordinate. We then introduce a simple phenomenological model of 1D FES with adjustable shape (symmetry, minima and barrier curvatures, and barrier height) and use it to simulate diffusive molecular trajectories coupled to stochastic donor/acceptor photon emissions. From these simulations, we develop some practical guidelines for the interpretation of transition paths and the analysis of photon trajectories with various maximum likelihood approaches. We find that, in principle, there is sufficient information on the photon trajectories to determine the shape of the free energy projected onto E (provided that E is a suitable reaction coordinate) as well as the intramolecular diffusion coefficient. We hence propose to directly perform the MLA of experimental photon trajectories with a diffusive 1D FES model of adjustable shape, which lends to more accurate TPT values and to mechanistic information than is not obtainable from discrete kinetic models.

METHODS

Landau Free Energy Surface Model. To represent a projected 1D free energy analytically, we use a phenomenological model based on a Landau quartic polynomial. The model is similar to the variable barrier model one of us co-developed for the analysis of differential scanning calorimetry

https://doi.org/10.1021/acs.jpcb.1c05401
J. Phys. Chem. B 2021, 125, 12413−12425

experiments,[59] but in this case, it is defined according to a generic order parameter (x) rather than the unfolding enthalpy. In this model, the Gibbs free energy as a function of the order parameter is defined at isostability conditions ($G_0 \Rightarrow G(\alpha_U) = G(\alpha_F)$) as

$$G_0(x) = -2\beta\left(\frac{x}{\alpha_U}\right)^2 + |\beta|\left(\frac{x}{\alpha_U}\right)^4 \quad \text{for } x \le 0$$

$$G_0(x) = -2\beta\left(\frac{x}{\alpha_F}\right)^2 + |\beta|\left(\frac{x}{\alpha_F}\right)^4 \quad \text{for } x > 0 \tag{3}$$

where α_U and α_F are the positions of the unfolded and folded (U and F) minima along the order parameter x, respectively, and x is defined with the condition that $\max(x) + |\min(x)| = 1$. β defines the free energy barrier (placed at $x = 0$) that separates states U and F. The sign of β determines whether the region between the two minima is convex (positive barrier) or concave (downhill: unimodal); when positive, its magnitude determines the height of the free energy barrier ($\Delta G^{\ddagger}$). The ratio $\phi = |\alpha_U|/(|\alpha_U| + \alpha_F)$ defines the asymmetry of the free energy surface, with $\varphi = 0.5$ indicating a perfectly symmetric surface (i.e., barrier halfway between U and F). Therefore, the model can produce bimodal−unimodal 1D free energy surfaces with shape (barrier height and curvatures of minima and barrier) controlled by these four parameters. To account for any changes in free energy as a function of (un)folding thermodynamic bias, we introduce the linear function $G_{bias}(x) = \lambda x$, where $\lambda \equiv G_{bias}(\max(x)) - G_{bias}(\min(x))$. The probability density as a function of the order parameter is hence:

$$p(x) = \frac{\exp(-(G_0(x) - \lambda x)/RT)}{\int \exp(-(G_0(x) - \lambda x)/RT)\, dx} \tag{4}$$

The kinetics of the process is described as diffusion on the free energy surface $G_0(x) - \lambda x$ as determined by the intramolecular diffusion coefficient D, which is assumed to be constant (i.e., independent of the order parameter). For the calculations presented in this study, we set $\lambda = 0$, so $\Delta G^{\ddagger}$ is the same in both directions (folding and unfolding).

Analytical Calculation of Mean Transition Path Times. To calculate the transition path time (TPT) from the 1D free energy surface (eq 3) and diffusion coefficient, we followed the treatment developed by Hummer.[54] This procedure starts with the definition of the transition boundaries (x_0 and x_1) and the calculation of the splitting probabilities:

$$\phi_U(x) = \frac{\int_x^{x_1} \exp(G(x')/RT)\, dx'}{\int_{x_0}^{x_1} \exp(G(x')/RT)\, dx'}$$

$$\phi_F(x) = \frac{\int_{x_0}^{x} \exp(G(x')/RT)\, dx'}{\int_{x_0}^{x_1} \exp(G(x')/RT)\, dx'} \tag{5}$$

which allow to calculate the mean length of the transition paths as

$$\langle t_{TP} \rangle = \frac{\int_{x_0}^{x_1} \exp(-G(x)/RT)\, \phi_U(x)\, \phi_F(x)\, dx \int_{x_0}^{x_1} \exp(G(x')/RT)\, dx'}{D} \tag{6}$$

Diffusive Kinetics on the 1D Free Energy Surface. To calculate the diffusive dynamics on the surface, we used a discretized version of the free energy surface and the following diffusive rate matrix:[57]

$$\mathbf{K} = \begin{pmatrix} -k_{1,2} & k_{2,1} & 0 & 0 & 0 \\ k_{1,2} & -(k_{2,1} + k_{2,3}) & k_{3,2} & \dots & 0 \\ 0 & k_{2,3} & \dots & k_{n-1,n-2} & 0 \\ 0 & \dots & k_{n-2,n-1} & -(k_{n-1,n-2} + k_{n-1,n}) & k_{n,n-1} \\ 0 & 0 & 0 & k_{n-1,n} & -k_{n,n-1} \end{pmatrix} \tag{7}$$

where the microscopic rates to go from the ith microstate to the previous ($i - 1$) and following ($i + 1$) microstates are given by

$$k_{i,i-1} = \frac{D}{2\Delta x^2}\left(\frac{p_{i-1}}{p_i} + 1\right) \tag{8a}$$

$$k_{i,i+1} = \frac{D}{2\Delta x^2}\left(\frac{p_{i+1}}{p_i} + 1\right) \tag{8b}$$

This rate matrix can be solved as an eigenvalue problem using standard methods. The calculation provides the sets of eigenvectors and eigenvalues of the system. The slowest nonzero eigenvalue represents the global equilibration rate. An example of 1D free energy surface (for a symmetric profile at isostability conditions and $\beta = 3$ kJ/mol) is given in Figure S1, together with the amplitudes (as a function of the order parameter) of the three slowest nonzero eigenvalues in the relaxation in response to an infinitesimal free energy perturbation.

Stochastic Diffusive Kinetic Simulations. We carried out stochastic diffusive kinetic simulations on the discretized version of the free energy surfaces. Particularly, we used a grid of 121 points, which provides sufficient resolution to smoothly reproduce the shape of the surfaces (even for the most asymmetric ones). The simulations were performed at constant time intervals with an algorithm and procedure explained before.[57] Briefly, we defined three possible outcomes starting from microstate i after an elapse of time Δt: moving forward to $i + 1$, moving backward to $i - 1$, or staying at i. According to this move set, the time-dependent probability of each of the microstates is obtained from the relationships

https://doi.org/10.1021/acs.jpcb.1c05401
J. Phys. Chem. B 2021, 125, 12413−12425

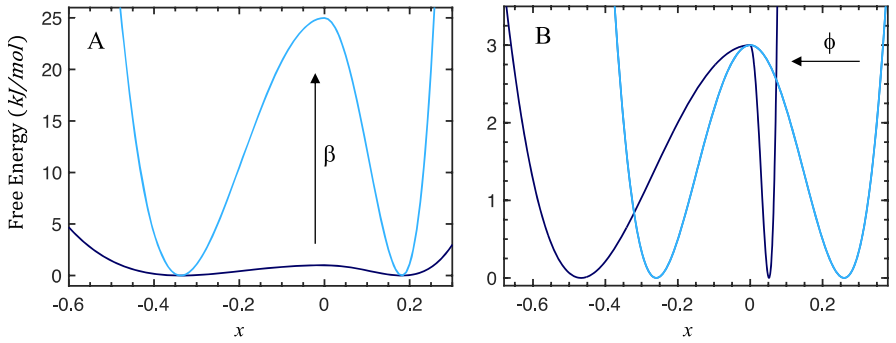


Figure 1. Landau 1D free energy surfaces used in this study. (Left) 1D free energy surfaces corresponding to the two extremes of the barrier range we used in this study. The light blue surface represents the highest barrier (10 k_BT), and the navy blue represents the lowest (0.4 k_BT). All these free energy surfaces are generated with $\phi = 0.65$. (Right) 1D free energy surfaces corresponding to the two extremes of the range we used to investigate the effects of the position of the barrier along the reaction coordinate. The light blue represents a surface with $\phi = 0.5$ (symmetric), and the navy blue represents a surface with $\phi = 0.9$ (highest asymmetry).

$$p(i \to i+1) = \Delta t \frac{D}{2\Delta x^2}\left(\frac{p_{i+1}}{p_i} + 1\right) \tag{9a}$$

$$p(i \to i-1) = \Delta t \frac{D}{2\Delta x^2}\left(\frac{p_{i-1}}{p_i} + 1\right) \tag{9b}$$

$$p(i \to i) = 1 - [p(i \to i-1) + p(i \to i+1)] \tag{9c}$$

when $\Delta t \ll (\Delta x)^2/D$. The time scale for moving along the surface is determined by the intramolecular diffusion coefficient D and the free energy gradient between neighboring microstates. Simulations with increasing barrier height were performed using proportionally faster D and smaller Δt to keep the total simulation time constant (20 s), while ensuring sufficient sampling statistics and enough number of transitions per burst on the photon trajectory simulations. The parameters used for each stochastic simulation are given in Table S1. Examples of diffusive molecular trajectories from simulations on surfaces of different barrier height are provided in Figure S2.

Analysis of Transition Paths from Stochastic Kinetic Trajectories. For condensed phase reactions, a transition path is the segment of a reactive diffusive trajectory ($U \to F$ or vice versa) that corresponds to the actual crossing of the barrier separating the two minima. Such a definition, of course, makes transition paths dependent on how one defines the effective transition region, which is somewhat arbitrary even when the free energy surface is known a priori. This problem is related to the better studied one of identifying the transition state ensemble (TSE). In this regard, the TSE is defined as the set of points in configuration space that share an equal probability of reaching the reactant region and the product region when propagated forward in time. An alternative definition explored by Hummer builds upon the concept of TSE regions as localities common to most reactive trajectories, that is, as those points in configuration space with the highest probability that equilibrium trajectories passing through them are indeed reactive.[54] Transition paths are then identified as trajectory segments that exit from a reactant region and reach a product region without crossing back into the initial reactant region, and vice versa. We used the Hummer approach to define and simulate transition paths. Namely, we define a transition path as a trajectory that starts from a point on the reactant basin of attraction that is on route to the barrier (x_0) and, when propagated forward in time, crosses over to an equivalent point on the product basin of attraction (x_1) without revisiting x_0. The transition paths in the opposite direction are simply defined by reversing x_0 and x_1. For a known 1D free energy surface, the TPT can be calculated analytically,[54] but it still requires that the boundaries of the transition region ($[x_0, x_1]$) are defined. We thus explored the effect on the TPT of the transition region boundaries while ensuring symmetry so that the relative distance from x_0 to the barrier top was equal on both sides (Figure 1). For each set of boundaries, we calculated the TPT analytically and also generated thousands of stochastic trajectories to obtain the distribution of transition path times in each direction (folding and unfolding) as well as the crossing probability. We define the crossing probability as the fraction of the total number of trajectory trials that starting at x_0 reached x_1 before returning to x_0. For comparative purposes, the transition paths for each simulation were rescaled by the ratio between a reference diffusion coefficient and the diffusion coefficient used in the specific simulation. Transition path times were then expressed relative to the transition path time obtained for the halfway boundaries and on the free energy surface with the highest barrier ($\phi = 0.65$ and 10 k_BT).

Simulated Time-Stamped Photon Trajectories. We simulated the outcome of single-molecule FRET experiments by performing stochastic simulations of donor/acceptor photon emission events as a function of time and the position of the molecule on the 1D free energy surface. The donor and acceptor emission rates were determined by the "experimental" count rate, which we set within realistic values for current SM-FRET experiments, and the probability of emitting a donor or an acceptor photon, as defined by the FRET efficiency (E) of each microstate on the surface ($p_A = E$ and $p_D = 1 - E$). For simulations with the lowest barriers, we used the same count-rate we obtained experimentally on the protein gpW, which folds/unfolds over a barrier of ~1 k_BT.[60] For simulations with higher barriers, we progressively increased the diffusion coefficient (D) to obtain approximately the same overall relaxation rate (similar numbers of folding/unfolding events in each simulated trajectory), and we adjusted Δt and photon count-rate accordingly to ensure comparable conditions (all the relevant parameters are given in Table S1).

We generated stochastic trajectories of interphoton times according to an exponential distribution with $\mu = 1/\mathrm{CR}$, where CR is the total photon (A + D) count rate, as explained before.[57] Each photon in the PT trajectory was then colored

https://doi.org/10.1021/acs.jpcb.1c05401
J. Phys. Chem. B 2021, 125, 12413−12425

(assigned an acceptor or donor tag) based on probabilities defined by the FRET efficiency of the microstate occupied by the molecule at the time the photon was emitted. The result of this procedure is a strip of time-stamped donor/acceptor photons emitted stochastically by the molecule as it is moving diffusively along the 1D free energy surface. This long strip of photons (4 to 25 million) was then divided into segments of 300−2000 photons to represent the number of photons per burst/trajectory that are typically obtained in SM-FRET experiments.

To determine the FRET efficiency as a function of the order parameter ($E(x)$), we used a linear function defined according to the empirical parameters E_U and E_F, corresponding to the E values at α_U (unfolded minimum) and α_F (folded minimum), respectively:

$$E(x) = E_U + \phi(E_F - E_U)\left(1 - \frac{x}{\alpha_U}\right) \quad \text{for } x \le 0$$
$$E(x) = E_{TS} + (1 - \phi)(E_F - E_U)\frac{x}{\alpha_F} \quad \text{for } x > 0 \tag{10}$$

where $E_{TS} = E_U + \phi(E_F - E_U)$ is the FRET efficiency at the barrier maximum ($x = 0$). For all the photon trajectory simulations in this work, we used $E_U = 0.574$ and $E_F = 0.885$, which are the E values we have obtained from the analysis of SM-FRET experiments on the fast folder gpW. These values are also in good agreement with those reported by other authors on single-domain proteins at their chemical denaturation midpoint.[51,52,61,62]

Maximum Likelihood Analysis of Mean Transition Path Times from Photon Trajectories. We used the Gopich−Szabo MLA-PT[49] to determine the TPT from the difference between the maximum likelihood of a two-state model (i.e., instantaneous transition: path not resolved) and that of a model with a virtual intermediate state (its lifetime provides the TPT).[51,52,63] The likelihood function for a photon trajectory with N photons is given by the expression

$$L_t = 1^T \prod_{j=2}^{N} [\mathbf{F}(c_j)\exp(\mathbf{K}\tau_j)]\,\mathbf{F}(c_1)\,p_{eq} \tag{11}$$

where $\mathbf{K}$ is the rate matrix for the given kinetic model, τ_j is the time interval between the jth and $(j-1)$th photon, p_{eq} is the vector with equilibrium probabilities of different states (as defined by $\mathbf{K}$), c_1 is the color of the first photon in the trajectory, and c_j is the color of the jth photon. $\mathbf{F}$(acceptor) = E and $\mathbf{F}$(donor) = $\mathbf{I} - E$, where E is a diagonal matrix with the FRET efficiencies of the states included in the model (ε_1, ..., ε_N). The total likelihood of a set of photon trajectories was maximized in base 10 logarithm units to avoid number overflow and minimize numerical precision errors. Hence, the total likelihood for an entire data set was calculated as $\log L = \sum_N \log L_t$, and then maximized. The procedure involves comparing the likelihoods of two- and three-state models. The fit to the two-state model is performed according to

$$U \rightleftarrows F$$
$$p_F + p_U = 1 \tag{12a}$$
$$k = k_f + k_u \tag{12b}$$
$$k_f p_U = k_u p_F \tag{12c}$$

where p_F is the fraction of protein in the native state and p_U is the fraction of protein in the unfolded state with FRET efficiencies, ε_F and ε_U, respectively. k_f and k_u are the folding and unfolding rate coefficients, which result in the rate matrix

$$\mathbf{K} = \begin{pmatrix} -k_f & k_u \\ k_f & -k_u \end{pmatrix} \tag{12d}$$

The three-state model represents a finite transition path time, defined by introducing an intermediate state S with lifetime $\tau_S = 1/2k_s$ that represents the average transition path time $\langle t_{TP}\rangle$. The kinetic scheme and parameters for the three-state model are

$$U \leftrightarrows S \rightleftarrows F$$
$$p_{F'} + p_S + p_{U'} = 1 \tag{13a}$$
$$k_f p_{U'} = k_u p_{F'} = p_S k_s \tag{13b}$$
$$\mathbf{K} = \begin{pmatrix} -k_f & k_s & 0 \\ k_f & -2k_s & k_u \\ 0 & k_s & -k_u \end{pmatrix} \tag{13c}$$

where $p_{U'}$, $p_{F'}$, and p_S are the fraction of protein molecules in the unfolded, folded, and intermediate states, respectively. $\Delta\log L(\tau_S) = \log L(S) - \log L(0)$ is calculated for a grid of τ_S and ε_S values, where $\log L(0)$ is the likelihood for the two-state model with an instantaneous transition and $\log L(S)$ is the likelihood for the three-state model with a finite transition. Each kinetic model was fit to the photon trajectory data using the MATLAB multidimensional unconstrained nonlinear minimization (Nelder−Mead) algorithm *fminsearch* by minimizing the negative of $\log L$. The two-state model was fit to obtain ε_F, ε_U, k_f, and k_u. For the three-state model (finite transition path), k_f and k_u were fixed to the values obtained with the two-state model. The FRET efficiencies for the three states were obtained in two ways: (a) ε_F and ε_U from the two-state model and ε_S fixed as halfway: This is a simpler, more common method of obtaining the TPT,[51] and will be referred to as $\Delta\log L_{2si\text{-}2s}$ in this work. (b) ε_F and ε_U obtained from the MLA-PT of the three-state model by fixing k_f and k_u to the values obtained from the two-state fit: $\Delta\log L(S)$ is then calculated for a 2D grid of values of ε_S (in the range $\varepsilon_U \le \varepsilon_S \le \varepsilon_F$) and τ_S. The combination of ε_S and τ_S values that gives the highest $\Delta\log L(\tau_S)$ is selected. We will be referring to this method as $\Delta\log L_{3s\text{-}2s}$ in this work.

These two methods are essentially equivalent with one difference: method (a) is simpler and assumes that the estimates of ε_F and ε_U are not affected by the introduction of the virtual intermediate state (i.e., ε_F and ε_U values from the two-state fit are correct); method (b) allows for relaxing ε_F and ε_U to accommodate for the presence of the virtual intermediate state. We introduced the latter because we realized that for surfaces with low free energy barriers the two-state fit tends to bring the two minima closer (moves ε_F and ε_U away from their true values and toward each other).

Maximum Likelihood Analysis of Photon Trajectories As Diffusion on a Free Energy Surface. To determine the ability to extract the free energy surface and dynamics of the system from photon trajectories, we analyzed them with the

https://doi.org/10.1021/acs.jpcb.1c05401
J. Phys. Chem. B 2021, 125, 12413−12425

MLA-PT[49,64] implemented with the Landau free energy surface model (as described above). For these calculations we followed exactly the procedure we previously developed for a 1D free energy surface model with restricted shape in which the asymmetry of the barrier is fixed.[57] Particularly, we globally analyzed the simulated photon trajectories for each given free energy surface to maximize the total likelihood. The fitting procedure involved finding the values of the 5 model parameters that maximize the likelihood: β, ϕ (λ = 0) for the free energy surface; E_U and E_F for the FRET efficiency; D for the dynamics.

RESULTS AND DISCUSSION

1D Free Energy Surface. We carried stochastic kinetic simulations on 1D free energy surfaces with different shapes using the Landau free energy surface model that we introduce here (see "Methods" section). In addition, we set all the free energy surfaces to isostability conditions (λ = 0) to ensure an identical barrier height for folding and unfolding and hence facilitate comparison. Particularly, we generated free energy surfaces with barriers from 10 k_BT down to 0.4 k_BT, which is a range that covers the barriers estimated for two-state single-domain proteins[32] and fast folders[33,65] (Figure 1A). These surfaces have exactly the same shape, with the barrier placed at 0.65 along the reaction coordinate (ϕ = 0.65), which is consistent with the experimental average ϕ-value of over 800 mutations on 24 single-domain proteins.[66] We also generated free energy surfaces with different levels of asymmetry (Figure 1B). The asymmetry range that we explore here (ϕ from 0.5 to 0.9) covers the range in beta-Tanford (β_T) values determined from the ratio of the slopes of the chevron plot for two-state folding proteins.[67,68]

Defining the Transition Region: Transition Path Times and Crossing Probabilities. The analysis of transition paths relies on the definition of the boundaries ($[x_0, x_1]$), which is somewhat arbitrary. A simple strategy is to set the boundaries halfway between the reactant minimum and the barrier top (Figure 2), but there are no particularly strong arguments for using such definition beyond convenience.

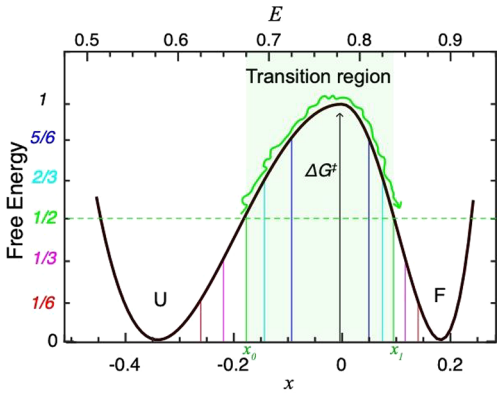


Figure 2. Transition region boundaries. The transition region is defined by its boundaries x_0 and x_1 on the reaction coordinate. Five examples are given in which x_0 and x_1 are defined as a fraction of the distance between the minimum and the barrier top (x = 0). The halfway (1/2) transition region is highlighted as a green swath with a schematic representation of a possible transition path as example. The reaction coordinate (x) is shown on the bottom, and the corresponding changes in E used for simulating photon trajectories are shown on the top.

However, the actual boundaries should have a significant impact on both the TPT and the crossing probability. We explored this issue analytically and via stochastic kinetic simulations using different boundaries (Figure 2) for all the free energy surfaces of this study. Figure 3 shows how the TPT (calculated analytically) and crossing probability change depending on the transition region boundaries for an exemplary free energy surface (ϕ = 0.65 and 1.2 k_BT). The results as a function of the barrier height are given in Figure S3.

As expected, Figure 3 shows that the TPT is strongly influenced by the transition region boundaries, changing by nearly 20-fold between a narrow region (5/6) and a wide region (1/6). The same trend was found on all other free energy surfaces (Figure S3). However, we also observed that the sensitivity to the transition region boundaries is less marked the higher is the barrier, being reduced to about 7-fold for the 10 k_BT barrier. This result is consistent with theoretical studies that found transition paths to be roughly insensitive to its precise definition for transitions that take place over barriers significantly higher than k_BT.[37] The crossing probability, which provides a measure of how often a molecule that starts at one boundary successfully crosses to the other, follows the inverse trend, increasing as the transition region becomes narrower. This is again expected as longer paths necessarily involve many more chances for the molecule to revert its trajectory. As the region becomes narrower, the crossing probability approaches 0.5, in agreement with the TSE (barrier top) in transition state theory.[69]

These results highlight the importance of the choice of boundaries for the analysis of transition paths, particularly for the modest free energy barriers expected for the (un)folding of single-domain proteins.[24] The question is which is definition is most useful. The analysis of the eigen spectra of the diffusive rate matrix offers one point of reference. As shown in Figure S1, the second slowest nonzero eigenvalue (λ_2) represents the equilibration between the barrier top and the two minima; hence, it is closely related to the reactive transition paths. Figure 3A shows that the inverse of this eigenvalue is very similar to the TPT for the halfway transition boundaries. However, λ_2 reflects the flux from either basin to the top rather than the crossing from one basin to the other. We hence conclude that $1/\lambda_2$ represents a lower bound for the effective TPT, and therefore, that the transition region should ideally be somewhat broader than halfway. An alternative, empirical way to address this question is to look for the transition region boundaries that produce the best scaling between the ratio $\tau_{kin}/(2\pi\langle\tau_{TP}\rangle)$ and $\Delta G^{\ddagger}$ over the range of barriers that is most significant for protein (un)folding.[52] Here τ_{kin} is the overall relaxation time ($1/\lambda_1$), and $\langle\tau_{TP}\rangle$ is the TPT for each boundary definition. This calculation for the data from Figure S2 shows again that the definition of transition region boundaries has small impact for the highest barriers. However, this calculation overestimates $\Delta G^{\ddagger}$ for the smaller barriers in general, and increasingly so as the boundaries of the transition region become narrower. For instance, the TPT for the narrowest transition region (5/6) results on $\Delta G^{\ddagger} \sim 3\ k_BT$ for the surface with a 0.4 k_BT barrier. Hence, we decided to use the 1/3 transition region as the most convenient definition for the analysis of transition paths in the range of free energy barriers that we explore here.

Effects of the Free Energy Barrier on the Transition Paths. The rate and TPT obtained from experiments have been used to estimate the barrier height by combining eqs 1

https://doi.org/10.1021/acs.jpcb.1c05401
J. Phys. Chem. B 2021, 125, 12413−12425

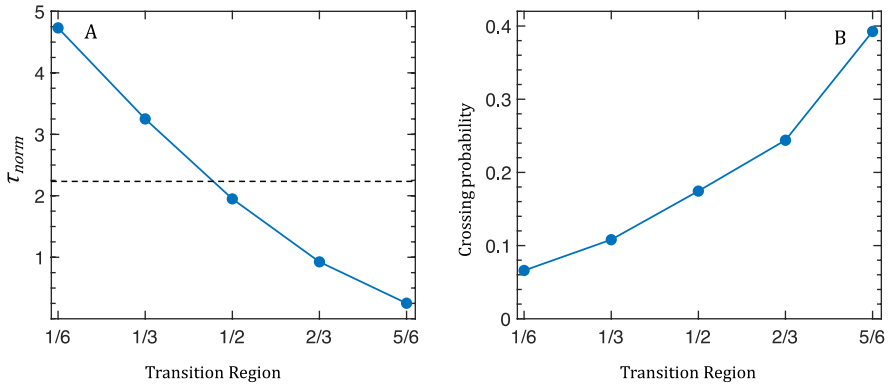


Figure 3. Transition path time (TPT) and crossing probability as a function of the transition region boundaries. (A) TPT calculated analytically for the different transition region boundaries (as in Figure 2) on a free energy surface with ϕ = 0.65 and 1.2 k_BT barrier. The dashed horizontal line indicates the inverse of the second slowest nonzero eigenvalue of the rate matrix ($1/\lambda_2$; see Figure S1). (B) Crossing probability.

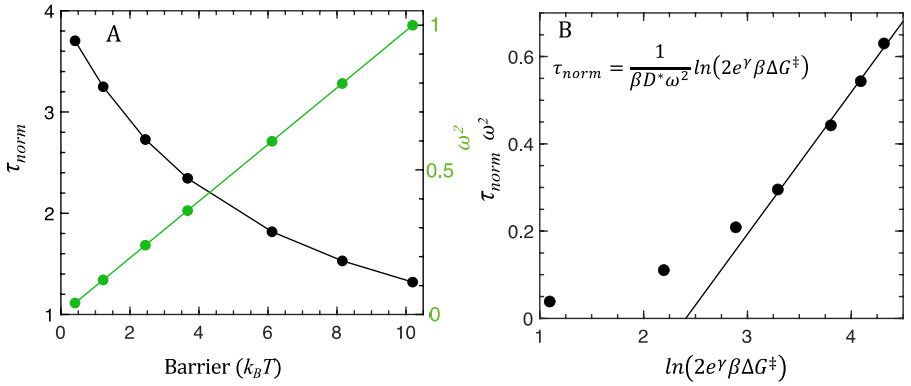


Figure 4. (A) TPT (black, left axis) and free energy barrier curvature (green, right axis) as a function of the barrier height. The barrier curvature is shown normalized with respect to the highest barrier used in this study. (B) Barrier dependence of the TPT multiplied by the barrier curvature. The line represents the expectation from the Szabo transition path time analytical expression.

and 2.[35] This procedure involves the assumption that the curvatures for both minima and barrier are equal and the diffusion coefficient is constant, so all these terms can be dropped from the ratio of eqs 1 and 2. Under those assumptions, the TPT is expected to increase proportionally to the barrier by $\ln[2e^{\gamma}\beta\Delta G^{\ddagger}]$. However, the TPT, determined either analytically or from stochastic kinetic simulations for free energy surfaces with the same shape but different barrier heights, exhibits the opposite trend; namely, the TPT increases monotonically as the barrier drops from 10 to 0.4 k_BT (Figure 4A). The effect is significant with a TPT about 3-fold longer for the marginal barrier than for the highest. The data in Figure 4 corresponds to the 1/3 transition region, but similar results are obtained for other transition boundaries (Figure S3). The analysis of the crossing probabilities sheds some light onto this seemingly counterintuitive result. The crossing probability, which reflects how many attempts to cross the barrier from x_0 are indeed reactive, decreases abruptly as the transition region becomes wider (Figure S3B). This is because longer paths have a higher probability for the molecule to reverse its trajectory at some point. Such effect is amplified by the gradient on the free energy surface; hence, the crossing probability drops dramatically when the barrier raises (about 3 orders of magnitude from 0.4 to 10 k_BT, see Figure S3B). In other words, a steeper free energy gradient enhances the probability of a trajectory reversal occurring during the climbing segment of the path, which makes most trials unproductive and slows down the rate. However, by the same token, the few trajectories that do indeed make it over the top are those that do not linger, that is, those that minimize the crossing time, which results in shorter TPTs.

The hidden factor behind the actual TPT barrier dependence is the curvature of the surface. The free energy surfaces used for this analysis have an identical shape: the same position of the minima and barrier top. However, as the barrier height increases, so does its curvature (green in Figure 4A), which affects the TPT in inversely proportional fashion, as shown in eq 2. We note that the second derivative of the Landau free energy surface is discontinuous at the barrier top (eq 3). Hence, we calculated the curvature separately on each side of the barrier top and averaged them out to obtain the barrier curvature. Figure 4B shows that when the TPT is multiplied by the barrier curvature of the corresponding surfaces the data closely follow eq 2 for barriers >2 k_BT (third point in Figure 4B). Only when the barrier becomes comparable to thermal energy does the TPT diverge from eq 2, as expected since this equation was derived for high barrier scenarios. Similar conclusions have been reached from theoretical analyses of the shape of transition paths.[36,37,53] Importantly, these results indicate that faster folding proteins are likely to exhibit longer transition paths, a conclusion of practical significance for the experimental analysis of protein (un)folding rates and transition paths. In other words, the best candidates for the single-molecule analysis of transition paths are fast-folding proteins, because their transition paths are longer and they undergo many more reactive transitions per unit of time.

https://doi.org/10.1021/acs.jpcb.1c05401
J. Phys. Chem. B 2021, 125, 12413−12425

Asymmetry of the Free Energy Surface and Transition Paths. Theoretical analysis indicates that it is possible to extract the shape and height of the free energy surface from the shape of the transition paths.[36,37,53] In contrast, existing methods for the analysis of SM-FRET experiments only produce TPT estimates; hence, there is a need for an independent measurement of the free energy surface.[35] Here we further explore the connection between the shape of the free energy surface and the transition path times using surfaces with different levels of asymmetry: from perfectly symmetric to a surface in which the barrier is at 90% of the distance between U and F (Figure 1B). We note here that in nonsymmetric surfaces, the transition state, defined as the position on the surface with splitting probabilities of 0.5, does not coincide with the barrier top. The transition state is in fact shifted toward the broader side of the barrier proportionally to the degree of asymmetry on the surface. This discrepancy is, however, of not much practical consequence for the calculation of the TPT, whether analytically with eqs 5 and 6 or from simulations. We thus generated transition paths in both folding and unfolding directions on surfaces with varying asymmetry (Figure 5). The TPT is the same in the forward (folding) and reverse (unfolding) directions, as expected from time reversal symmetry (blue on top and middle panels of Figure 5). Moreover, in general, the TPT is not very sensitive to the asymmetry of the free energy surface, as it has been theoretically proposed.[36,37,53] This is particularly true for small barriers. For instance, on the lowest barrier of this study (~0.4 k_BT), the TPT increases by just ~4% from the symmetric to a ϕ = 0.9 surface, whereas the TPT increases by ~36% on the largest barrier (~ 10 k_BT). The longer TPT is due to the fact that as the surface grows in asymmetry the segment of the transition path that corresponds to the broad side of the barrier becomes an increasingly larger fraction of the total path. Since the free energy gradient on that segment of the path is shallower, the overall crossing time increases. This effect is equivalent to the TPT becoming longer as the barrier decreases (Figure 4A), which also explains why surfaces with lower barriers are much less sensitive to this effect.

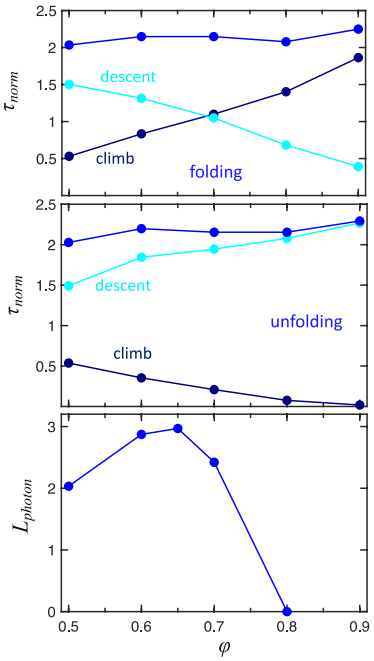


Figure 5. TPT as a function of the asymmetry of the free energy surface (ϕ). (Top) TPT in the folding direction ($U \rightarrow F$), showing the climb (navy) and descent (cyan) segments of the path. (Middle) As shown in the top panel but in the unfolding direction ($F \rightarrow U$). (Bottom) Likelihood per photon (multiplied by 10^4) from the MLA of photon trajectories simulated on a ϕ = 0.65 surface as obtained using Landau free energy surfaces at fixed asymmetry values.

That observation also points to the convenience of separating the path into climb and descent segments as a way to extract information about the surface's asymmetry. The climb corresponds to the segment from the moment the molecule enters the transition region (x_0) to the instant that it reaches the barrier top ($x = 0$) for the first time. The descent corresponds to the time it takes to reach x_1 for the first time starting from $x = 0$, but the molecule can return to $x = 0$ multiple times as long as it does not backtrack all the way to x_0. These segments are equivalent to $t_{TP}(x = 0 \mid x_0)$ and $t_{FP}(x_1 \mid x = 0)$ in Makarov's definition,[37] respectively. These two segments are identical for folding and unfolding transitions paths, but in reverse: swapping x_0 and x_1. The separation of TPT into climb and descent segment does indeed reveal interesting symmetry-dependent properties (Figure 5). For the symmetric barrier case, we note that the downhill descent takes about 3 times longer than the climb. This is again because the productive climbs minimize the path (or else they return to the originating basin and are not considered), whereas the descent allows for much more lingering. However, the most important result is that the climb and descent TPT segments are very sensitive to the free energy surface's asymmetry (Figure 5). The implication is that the ratio between climb and descent TPT segments provides key information about the shape of the free energy surface for nontrivial cases in which the barrier curvature is not symmetric (hence, the rate is not exactly described by eq 1, and the transition state does not coincide with the barrier top). This finding has potentially great significance for the single-molecule characterization of protein (un)folding reactions. The reason is that obtaining estimates of the average time for each segment is much easier than fully resolving the shape of the path. We further explore this issue in the following sections.

Estimating the Mean Transition Path Time from Photon Trajectories. MLA-PT is the approach available to estimate the TPT from single-molecule FRET experiments. One of its limitations is the need a priori of a kinetic model to compute and maximize the likelihood.[51,52,63] In its simplest version, this procedure compares the likelihoods obtained with a two-state and a simplified three-state model with a virtual intermediate. The lifetime of the virtual intermediate in the three-state model represents the TPT; hence, the comparison between the three-state likelihood as a function of the intermediate lifetime ($1/2k_S$) and the likelihood of the two-state model ($\Delta\log L_{2si\text{-}2s}$) provides an estimate of whether the TPT is resolved and of its value (the maximum along the $\Delta\log L_{2si\text{-}2s}$ curve). This methodology has been recently tested by Taumoefolau and Best using coarse-grained molecular simulations of protein folding to explore fundamental issues such as the significance of using projected surfaces and E as reaction coordinate.[58] Here we perform a similar test but using photon trajectories generated from the diffusive stochastic kinetics on 1D free energy surfaces with E as reaction coordinate. This type of analysis informs us of the inherent accuracy and limitations of the approach for an idealized

https://doi.org/10.1021/acs.jpcb.1c05401
J. Phys. Chem. B 2021, 125, 12413−12425

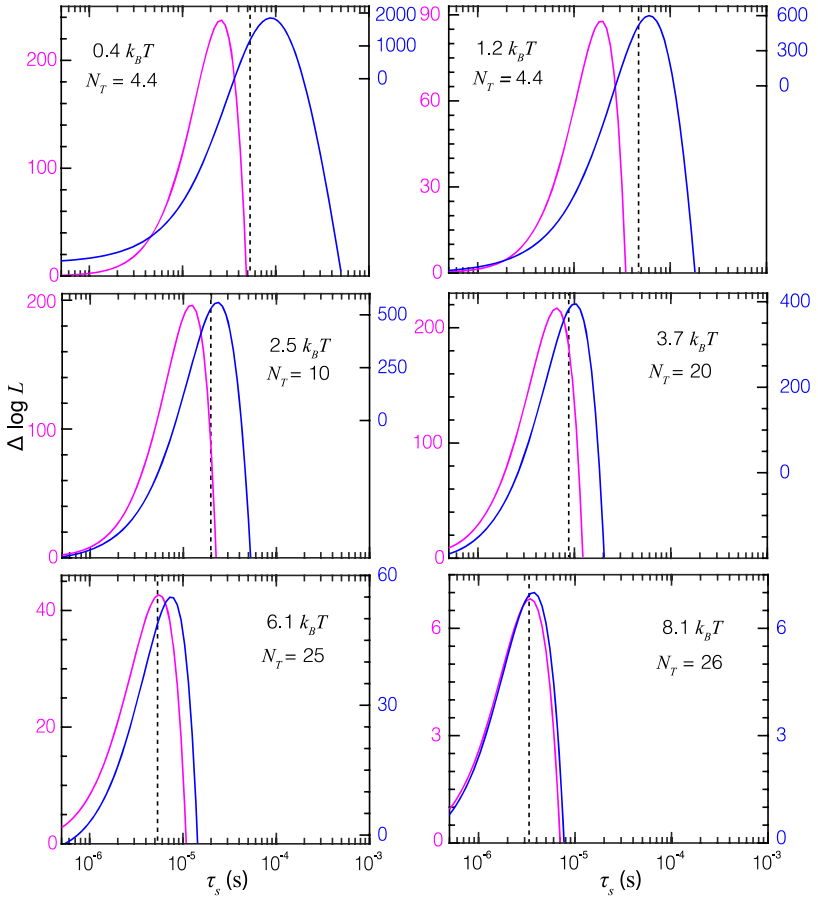


Figure 6. Δlog L plots obtained from photon trajectory simulations on free energy surfaces with varying barrier (simulation parameters in Table S1). Results for simulations with increasing barrier are shown from top left to bottom right with the barrier height indicated in units of k_BT: Δlog $L_{2s\text{-}2si}$ (magenta, left scale) and Δlog $L_{2s\text{-}3s}$ (blue, right scale) methods. The vertical dashed line signals the actual TPT using 1/3 boundaries for the transition region. N_T is the number of million photons used in the analysis.

problem; thus, it is complementary to that of Taumoefolau and Best. It also provides a direct point of reference for the analysis of experiments. For this purpose, we generated stochastic diffusive kinetic trajectories on 1D free energy surfaces with asymmetry of ϕ = 0.65 and barriers ranging from 0.4 to 8.1 k_BT (Figure 1A). We then simulated photon emission trajectories associated with the intrinsic dynamics on the surface using a simple linear function $E(x)$ defined based on the E_U and E_F values (E values at the two minima), which we chose to be consistent with typical SM-FRET experiments of single-domain protein (un)folding at the chemical denaturation midpoint. We then performed the two-/three-state MLA-PT as described above and in the "Methods" section. The results of these analyses are summarized in Figure 6.

The first point to note is that although log L is proportional to the total number of photons included in the analysis, Δlog L is only sensitive to the fraction of photons emitted during the actual transition paths. In the original implementation, the analysis is only performed on the small segments of photon trajectories that contain a transition (identified independently using a heuristic criterion, such as the detection of a sharp change in E in the binned photon data). This also requires that the ε_U and ε_F parameters are determined independently and fixed during the procedure,[51,52] because the short trajectory segments including the transition paths do not have enough information to define the positions of the minima. We tried this procedure using different heuristic criteria to identify transitions and procedures to estimate ε_U and ε_F. We also performed a direct analysis of all the photon data without preselection and globally fitting ε_U, ε_F, k_u, and k_f with the two-state model (see the "Methods" section). The latter procedure is more straightforward because it eliminates the need for identifying the specific segments corresponding to transitions, which is not easy to do when the transition path times are just slightly longer than the interphoton times, $E_U \sim E_F$, or there is some degree of acceptor blinking/bleaching.[70] The only issue against performing a global analysis is the added computational cost. However, the analysis is trivially parallelizable by distributing photon trajectories among multiple threads/cores, so we strongly recommend performing this analysis using all the available photons, especially for fast-folding proteins with low barriers.

As for its performance, we find that Δlog $L_{2si\text{-}2s}$ reliably produces a maximum for all the free energy surfaces, at least up to a barrier of ~8 k_BT (magenta in Figure 6). Therefore, this procedure does indeed detect transition paths, even when they are only ~5 times longer than the average interphoton time (compare Figure 6 with count-rates in Table S1). For the highest barrier, the position of the Δlog $L_{2si\text{-}2s}$ maximum is nearly identical to the actual TPT obtained analytically for the 1/3 boundaries. As the barrier decreases, however, the estimated and actual TPT diverge, with the Δlog $L_{2si\text{-}2s}$ maximum underestimating the TPT by up to 2.5-fold for the lowest barriers (Figure 6). This result is consistent with the

https://doi.org/10.1021/acs.jpcb.1c05401
J. Phys. Chem. B 2021, 125, 12413−12425

findings of Taumoefolau and Best on surfaces projected from coarse-grained simulations, which had barriers on the lower end of the range we explore here.[58] Notably, the Δlog $L_{2si\text{-}2s}$ value at maximum follows the opposite trend, with the marginal 0.4 k_BT barrier producing a maximum at ∼240 (statistical significance of $1-10^{-240}$), whereas it is only ∼7 (significance of $1-10^{-7}$) for the 8.1 k_BT barrier (magenta scale in Figure 6). The difference in Δlog $L_{2si\text{-}2s}$ is even more drastic when it is corrected by the number of photons used in the analysis (Table 1). However, we note that this difference is

Table 1. Difference in Likelihood Values per Photon (Expressed in Log Units) for Different Methods

$\Delta G^{\ddagger}$ (k_BT)	Δlog $L_{FES\text{-}2s}$	Δlog $L_{3s\text{-}2s}$	Δlog $L_{2si\text{-}2s}$
0.4	4.76×10^{-4}	4.18×10^{-4}	5.4×10^{-5}
1.2	1.76×10^{-4}	1.36×10^{-4}	1.99×10^{-5}
2.5	7.38×10^{-5}	5.6×10^{-5}	1.96×10^{-5}
3.7	3.02×10^{-5}	2.0×10^{-5}	1.08×10^{-5}
6.1	5.21×10^{-6}	2.2×10^{-6}	1.7×10^{-6}
8.2	7.72×10^{-7}	2.7×10^{-7}	3.0×10^{-9}

simply a reflection of the number of "significant" photons that are contained in each data set, that is, the photons that are emitted during the transition paths. A simple way to correct for the number of "significant" photons is to multiply Δlog $L_{2si\text{-}2s}$ per photon (Table 1) by $\exp(\Delta G^{\ddagger})$. Such a calculation shows that the Δlog $L_{2si\text{-}2s}$ per "significant" photon is almost 5-fold higher for the tallest relative to the shallowest barriers, explaining why this method performs well on two-state-like free energy surfaces, but underestimates the TPT when the barrier drops below 4 k_BT (Figure 6).

We also tried an alternative procedure in which the analysis with the three-state model is performed with fixed k_u and k_f (values from the two-state analysis), but ε_U and ε_F are adjusted together with k_S to accommodate for the virtual intermediate (Δlog $L_{3s\text{-}2s}$). This analysis generally produced much higher Δlog L maxima (blue in Figure 6), increasingly so as the barrier lowers. This is expected since lower barriers are more populated and hence more consistent with the presence of a virtual intermediate. However, the Δlog $L_{3s\text{-}2s}$ as a function of the intermediate lifetime lends a broader curve with its maximum overestimating the TPT, particularly for the intermediate to low barriers. The most important practical result from this comparative analysis is that, together, the two procedures set a useful range and their mean provides a good estimate of the actual TPT even when the barrier is very small (Figure 7).

Maximum Likelihood Analysis as Diffusion on a 1D Free Energy Surface. The combination of Δlog $L_{2si\text{-}2s}$ and Δlog $L_{3s\text{-}2s}$ can thus lead to fairly accurate TPT estimates. However, they do not provide any information about the underlying free energy surface, nor an accurate estimate of the barrier height given that barrier curvature and height are directly linked (Figure 4). In principle, if the rate of a process can be effectively described as diffusion on a free energy surface, then the photon trajectories from single molecules should contain all the information about the free energy surface and diffusion coefficient. The key question is whether, or to what extent, such information can be extracted via MLA-PT. This question has been addressed for idealized 1D free energy surfaces in which the barrier height varied but the specific shape of the surface was kept constant.[57] Here we aim

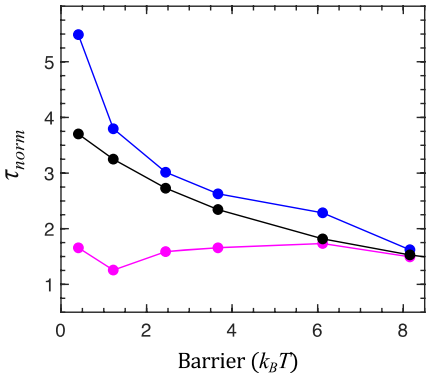


Figure 7. Relative TPT (normalized as before) as a function of the free energy barrier. The normalized actual TPT is shown in black. The TPT estimated from the maximum of the Δlog $L_{2s\text{-}2si}$ and Δlog $L_{2s\text{-}3s}$ plots of Figure 6 are shown in magenta and blue, respectively.

to determine whether it is practically feasible to extract the basic terms of a rate expression such as eq 1 (barrier height, curvatures of minima and barrier, and diffusion coefficient) by analyzing photon trajectories directly as diffusion on an adjustable 1D free energy surface (FES-MLA, see the "Methods" section).

The first observation from the FES-MLA is that the photon trajectory data that we used for the analysis with two-/three-state models (previous section) contain sufficient information to accurately retrieve the original free energy surface, regardless of how high is the barrier that separates the two minima (Figure S5). Δlog $L_{FES\text{-}2s}$ is in fact much higher than Δlog $L_{2si\text{-}2s}$, or even Δlog $L_{3s\text{-}2s}$, for all the barriers (Table 1). This result confirms that there is enough information in the photon trajectories to determine the barrier height and dynamics. Moreover, the FES-MLA also resolves the climb and descent segments of transition paths, and hence, it can determine the degree of asymmetry (shape) of the free energy surface (Figure 5, bottom). The comparison between Δlog $L_{FES\text{-}2s}$ and Δlog $L_{3s\text{-}2s}$ provides an empirical reference to ascertain whether a given photon data set can define the shape of the free energy surface relative to the presence of a discrete intermediate (the Landau MLA has the same number of fitting parameters as a restricted three-state model with single k_S; see the "Methods" section). This comparison can be particularly useful to discriminate between dynamics on a free energy surface with a shallow barrier and the relaxation of an intermediate separated from U and F by barriers. For instance, the Δlog L results summarized in Table 1 indicate that for our simulated conditions (protein dynamics, count-rates and E values) about 120 000 photons are sufficient to distinguish (with statistical significance of $1-10^{-12}$) the dynamics on a 0.4 k_BT surface from a relaxation through a discrete intermediate. Table 1 shows that such a shallow barrier is the most challenging scenario to discriminate, whereas resolving the diffusive transition paths becomes comparatively easier the higher is the barrier (provided the count rate is at least 4−5-fold higher than 1/TPT).

The comparative analysis reveals another byproduct of fitting the photon trajectories to discrete kinetic models; namely, we find that the two-state model does not place ε_F and ε_U where the surface minima are but rather closer together (magenta lines in Figure S5). The discrepancy is minimal for the scenarios with the highest barriers, but it grows proportionally to the barrier shallowness. This effect seems

https://doi.org/10.1021/acs.jpcb.1c05401
J. Phys. Chem. B 2021, 125, 12413−12425

to reflect dynamic averaging with the population of the microstates at the barrier. To test this idea, we performed the two-state fitting with the actual ε_F and ε_U fixed to the known E_F and E_U values, which returned many orders of magnitude lower likelihoods, indicating that the effect does indeed reflect dynamic averaging from excursions along the barrier slope (for a 0.4 k_BT barrier the population at the top is 2/3 of that at the minimum). Because the shift in ε_F and ε_U reflects the true dynamics of the system, it is preferable to fit these parameters jointly with k_u and k_f, as we suggested above, rather than fixing them to expected values from an independent measurement. In fact, the observation of a shift in ε_F and ε_U between the two- and three-state analyses could be used as first indication that the TPT might be resolvable. The three-state model partially corrects the shift in ε_F and ε_U by introducing an intermediate. However, the three-state model fails to place ε_S at the exact barrier top: ε_S is shifted toward U for the lower barriers and toward F for the highest ones (blue lines in Figure S5). We note that these issues arise from the use of discrete kinetic models. Hence, we conclude that an MLA procedure that directly uses a model of diffusion on a simple 1D free surface is preferable, as it has been proposed before by us[57] and more recently by others.[58] This is particularly true if the shape of the idealized 1D surface (barrier height and asymmetry) can be controlled with specific parameters, as with the Landau model we introduce here.

CONCLUSIONS

There has been a growing interest in the analysis of the reactive transition paths of protein (un)folding[35] and other biomolecular reactions driven by large conformational changes.[55,56] This is so because such transition paths contain essential information to unravel the complex mechanisms that determine the rates of these processes.[23] A major motivating factor has been the recent developments in single-molecule fluorescence and force techniques, which have opened the opportunity to address these questions experimentally for the very first time. Here we tackled two related questions: (i) how to interpret transition paths in mechanistic terms and (ii) how to extract transition path information from single-molecule experiments. We started by assuming that the rates and transition paths of protein (un)folding and related processes can be described as diffusion on a simple free energy surface that represents the projection of the hyperdimensional energy landscape onto a single reaction coordinate. We describe such free energy surface analytically using a Landau quartic polynomial, which produces an idealized 1D surface with two minima separated by a barrier of controllable height and symmetry. Using stochastic diffusive simulations on such surfaces, we closely looked into the direct relationships between the shape of the free surface and the transition paths (shape and time). From this analysis we can extract three main conclusions: (a) The choice of transition region boundaries has strong impact on the resulting TPT and crossing probability. However, we find that setting the boundaries at 1/3 of the distance between the minimum and the barrier top best captures the dynamics on the surface, particularly for barriers in the range that is most relevant to protein folding reactions (0−10 k_BT). (b) For free energy surfaces with the same shape (symmetry), the TPT is inversely proportional to the barrier height, being the longest when the surface has a positive, marginally high barrier. We show that this behavior reflects the tight connection between barrier curvature and height, and it is consistent with the known expression $\langle \tau_{TP} \rangle \approx \ln [2e^{\gamma}\beta\Delta G^{\ddagger}]/D\beta(\omega^{\ddagger})^2$. Practically, this result implies that resolving the transition paths of fast-folding proteins should be easier because they will be both longer and much more frequent. (c) Whereas the TPT is insensitive to the surface asymmetry, as indicated by theoretical analysis,[36,37,53] we find that the climb and descent segments of the transition path are highly sensitive to the barrier's asymmetry. Therefore, resolving the climb and descent fractions of the TPT could provide key information to estimate the shape of the surface.

We then investigated the performance of various MLA-PT approaches for extracting transition path information. Our results show that the original analysis ($\Delta\log L_{2si\text{-}2s}$) underestimates the TPT, particularly for barriers smaller than 4 k_BT, consistently with recent results from a similar study based on coarse-grained molecular simulations.[58] The major factor behind such underestimation is that the two-state model shifts the two minima closer together in response to the population at the barrier, which makes the estimated transition path shorter than the real one. We introduce here a variant in which the three-state model is allowed to accommodate the position of the two minima ($\Delta\log L_{3s\text{-}2s}$). This variant overestimates the lifetime of the intermediate, and hence the TPT. However, the two MLA procedures combined provide a rather accurate mean TPT over the entire range of barriers explored here.

However, we do find that the MLA implemented with a model of diffusion on a free energy surface (MLA-FES) offers a far superior performance. Such analysis renders much higher likelihoods than the discrete kinetic models with equal number of parameters, indicating that there is sufficient information in the simulated photon trajectories to extract the diffusive dynamics on the free energy surface. We then confirm that this analysis obtains the correct barrier throughout the 0−10 k_BT range, as well as the asymmetry of the free energy surface. For a real experiment, in which the free energy surface is not known a priori, the researcher can compare $\Delta\log L_{FES\text{-}2s}$ and $\Delta\log L_{3s\text{-}2s}$ as indicator of the statistical significance of the FES analysis (e.g., Table 1). Accordingly, we strongly recommend that the analysis is performed with a flexible 1D free energy surface model, like the one we introduce here. This model is parametrically as simple as a three-state model, and it allows us to dissect the terms of the rate expression. Such representation is amenable to direct comparison with theory and the results from molecular simulations. Furthermore, the MLA-FES can easily accommodate more complex scenarios: a diffusion coefficient that changes along the reaction coordinate, as it has been postulated by theory[1] and observed in simulations,[18,71] or the use of higher dimensional surfaces to reproduce the results from experiments that monitor several distances in the protein (multiparametric and/or three-color FRET experiments).

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.1c05401.

Parameters used for the stochastic kinetic simulations, symmetric free energy surface and three slowest eigenvalues, examples of stochastic kinetic trajectories, TPTs and crossing probabilities, TPT dependence on barrier height, free energy surfaces and mean E values estimated by the different MLA methods described in the manuscript (PDF)

https://doi.org/10.1021/acs.jpcb.1c05401
J. Phys. Chem. B 2021, 125, 12413−12425

AUTHOR INFORMATION

Corresponding Author

Victor Muñoz − *NSF-CREST Center for Cellular and Biomolecular Machines (CCBM), Chemistry and Chemical Biology Graduate Program, and Department of Bioengineering, University of California, Merced 95343 California, United States;* orcid.org/0000-0002-5683-1482; Email: vmunoz3@ucmerced.edu

Author

Nivin Mothi − *NSF-CREST Center for Cellular and Biomolecular Machines (CCBM) and Chemistry and Chemical Biology Graduate Program, University of California, Merced 95343 California, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.1c05401

Author Contributions

The manuscript was written through contributions of both authors, who have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Science foundation (NSF-MCB-1616759) and the CREST Center for Cellular and Biomolecular Machines (grant NSF-CREST-1547848). V.M. acknowledges additional support from the W.M. Keck Foundation.

ABBREVIATIONS

MLA-PT	maximum likelihood analysis of photon trajectories
FES	free energy surface
TPT	mean transition path time
TSE	transition state ensemble
FRET	Förster resonance energy transfer
SM-FRET	single-molecule Förster resonance energy transfer
FS	force spectroscopy.

REFERENCES

(1) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct., Funct., Genet.* **1995**, *21* (3), 167−95.

(2) Onuchic, J. N.; Wang, J.; Wolynes, P. G. Analyzing single molecule trajectories on complex energy landscapes using replica correlation functions. *Chem. Phys.* **1999**, *247* (1), 175−184.

(3) Leite, V. B. P.; Onuchic, J. N.; Stell, G.; Wang, J. Probing the kinetics of single molecule protein folding. *Biophys. J.* **2004**, *87* (6), 3633−3641.

(4) Camacho, C. J.; Thirumalai, D. Kinetics and thermodynamics of folding in model proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (13), 6369−6372.

(5) Knott, M.; Chan, H. S. Exploring the effects of hydrogen bonding and hydrophobic interactions on the foldability and cooperativity of helical proteins using a simplified atomic model. *Chem. Phys.* **2004**, *307* (2−3), 187−199.

(6) Lammert, H.; Wolynes, P. G.; Onuchic, J. N. The role of atomic level steric effects and attractive forces in protein folding. *Proteins: Struct., Funct., Genet.* **2012**, *80*, 362−373.

(7) Guo, Z.; Thirumalai, D. The nucleation-collapse mechanism in protein folding: evidence for the non-uniqueness of the folding nucleus. *Folding Des.* **1997**, *2* (6), 377−391.

(8) Garcia, A. E.; Onuchic, J. N. Folding a protein in a computer: An atomic description of the folding/unfolding of protein A. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (24), 13898−13903.

(9) Bowman, G. R.; Pande, V. S. Protein folded states are kinetic hubs. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (24), 10890−10895.

(10) Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (44), 17874−17879.

(11) Best, R. B. Atomistic molecular simulations of protein folding. *Curr. Opin. Struct. Biol.* **2012**, *22* (1), 52−61.

(12) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; et al. Atomic-level characterization of the structural dynamics of proteins. *Science* **2010**, *330* (6002), 341−346.

(13) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334* (6055), 517−20.

(14) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Protein folding kinetics and thermodynamics from atomistic simulation. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (44), 17845−17850.

(15) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. Diffusive dynamics of the reaction coordinate for protein folding funnels. *J. Chem. Phys.* **1996**, *104* (15), 5860−5868.

(16) Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940**, *7* (4), 284−304.

(17) Nelson Onuchic, J.; Nymeyer, H.; Garcia, A. E.; Chahine, J.; Socci, N. D. The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Adv. Protein Chem.* **2000**, *53*, 87−152.

(18) Best, R. B.; Hummer, G. Coordinate-dependent diffusion in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (3), 1088−1093.

(19) Li, M. S.; Klimov, D. K.; Thirumalai, D. Thermal denaturation and folding rates of single domain proteins: size matters. *Polymer* **2004**, *45* (2), 573−579.

(20) Plotkin, S. S.; Onuchic, J. N. Understanding protein folding with energy landscape theory. Part II: Quantitative aspects. *Q. Rev. Biophys.* **2002**, *35* (3), 205−86.

(21) Best, R. B.; Hummer, G. Diffusion models of protein folding. *Phys. Chem. Chem. Phys.* **2011**, *13* (38), 16902−16911.

(22) Makarov, D. E. Reconciling transition path time and rate measurements in reactions with large entropic barriers. *J. Chem. Phys.* **2017**, *146* (7), No. 071101.

(23) Satija, R.; Berezhkovskii, A. M.; Makarov, D. E. Broad distributions of transition-path times are fingerprints of multi-dimensionality of the underlying free energy landscapes. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (44), 27116−27123.

(24) Campos, L. A.; Sadqi, M.; Muñoz, V. Lessons about Protein Folding and Binding from Archetypal Folds. *Acc. Chem. Res.* **2020**, *53* (10), 2180−2188.

(25) Prigozhin, M. B.; Gruebele, M. Microsecond folding experiments and simulations: a match is made. *Phys. Chem. Chem. Phys.* **2013**, *15* (10), 3372−3388.

(26) Muñoz, V.; Cerminara, M. When fast is better: protein folding fundamentals and mechanisms from ultrafast approaches. *Biochem. J.* **2016**, *473* (17), 2545−59.

(27) Eaton, W. A.; Muñoz, V.; Thompson, P. A.; Henry, E. R.; Hofrichter, J. Kinetics and Dynamics of Loops, α-Helices, β-Hairpins, and Fast-Folding Proteins. *Acc. Chem. Res.* **1998**, *31* (11), 745−753.

(28) Sadqi, M.; Lapidus, L. J.; Muñoz, V. How fast is protein hydrophobic collapse? *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (21), 12117−22.

(29) Yang, W. Y.; Gruebele, M. Folding at the speed limit. *Nature* **2003**, *423* (6936), 193−7.

(30) Li, P.; Oliva, F. Y.; Naganathan, A. N.; Muñoz, V. Dynamics of one-state downhill protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (1), 103−8.

(31) Kubelka, J.; Hofrichter, J.; Eaton, W. A. The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* **2004**, *14* (1), 76−88.

(32) Akmal, A.; Muñoz, V. The nature of the free energy barriers to two-state folding. *Proteins: Struct., Funct., Genet.* **2004**, *57* (1), 142−52.
(33) Naganathan, A. N.; Doshi, U.; Muñoz, V. Protein folding kinetics: barrier effects in chemical and thermal denaturation experiments. *J. Am. Chem. Soc.* **2007**, *129* (17), 5673−82.
(34) Szczepaniak, M.; Iglesias-Bexiga, M.; Cerminara, M.; Sadqi, M.; Sanchez de Medina, C.; Martinez, J. C.; Luque, I.; Munoz, V. Ultrafast folding kinetics of WW domains reveal how the amino acid sequence determines the speed limit to protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (17), 8137−8142.
(35) Chung, H. S.; Eaton, W. A. Protein folding transition path times from single molecule FRET. *Curr. Opin. Struct. Biol.* **2018**, *48*, 30−39.
(36) Chaudhury, S.; Makarov, D. E. A harmonic transition state approximation for the duration of reactive events in complex molecular rearrangements. *J. Chem. Phys.* **2010**, *133* (3), No. 034118.
(37) Makarov, D. E. Shapes of dominant transition paths from single-molecule force spectroscopy. *J. Chem. Phys.* **2015**, *143* (19), 194103.
(38) Chung, H. S.; Louis, J. M.; Eaton, W. A. 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.* **2009**, *106* (29), 11837−11844.
(39) Wang, Z.; Campos, L. A.; Muñoz, V. Single-Molecule Fluorescence Studies of Fast Protein Folding. *Methods Enzymol.* **2016**, *581*, 417−459.
(40) Carrion-Vazquez, M.; Oberhauser, A. F.; Fisher, T. E.; Marszalek, P. E.; Li, H.; Fernandez, J. M. Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Prog. Biophys. Mol. Biol.* **2000**, *74* (1−2), 63−91.
(41) Guinn, E. J.; Jagannathan, B.; Marqusee, S. Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein. *Nat. Commun.* **2015**, *6* (1), 6861.
(42) Zhuravlev, P. I.; Hinczewski, M.; Chakrabarti, S.; Marqusee, S.; Thirumalai, D. Force-dependent switch in protein unfolding pathways and transition-state movements. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (6), E715−E724.
(43) Schonfelder, J.; Perez-Jimenez, R.; Muñoz, V. A simple two-state protein unfolds mechanically via multiple heterogeneous pathways at single-molecule resolution. *Nat. Commun.* **2016**, *7*, 11777.
(44) Yu, H.; Gupta, A. N.; Liu, X.; Neupane, K.; Brigley, A. M.; Sosova, I.; Woodside, M. T. 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.* **2012**, *109* (36), 14452−14457.
(45) Neupane, K.; Foster, D. A.; Dee, D. R.; Yu, H.; Wang, F.; Woodside, M. T. Direct observation of transition paths during the folding of proteins and nucleic acids. *Science* **2016**, *352* (6282), 239−242.
(46) Schönfelder, J.; De Sancho, D.; Berkovich, R.; Best, R. B.; Muñoz, V.; Perez-Jimenez, R. Reversible two-state folding of the ultrafast protein gpW under mechanical force. *Communications Chemistry* **2018**, *1* (1), 59.
(47) Cossio, P.; Hummer, G.; Szabo, A. On artifacts in single-molecule force spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (46), 14248−14253.
(48) De Sancho, D.; Schonfelder, J.; Best, R. B.; Perez-Jimenez, R.; Muñoz, V. Instrumental Effects in the Dynamics of an Ultrafast Folding Protein under Mechanical Force. *J. Phys. Chem. B* **2018**, *122* (49), 11147−11154.
(49) Gopich, I. V.; Szabo, A. Decoding the pattern of photon colors in single-molecule FRET. *J. Phys. Chem. B* **2009**, *113* (31), 10965−73.
(50) Chung, H. S.; Cellmer, T.; Louis, J. M.; Eaton, W. A. Measuring ultrafast protein folding rates from photon-by-photon analysis of single molecule fluorescence trajectories. *Chem. Phys.* **2013**, *422*, 229−237.
(51) Chung, H. S.; McHale, K.; Louis, J. M.; Eaton, W. A. Single-molecule fluorescence experiments determine protein folding transition path times. *Science* **2012**, *335* (6071), 981−4.
(52) Chung, H. S.; Eaton, W. A. Single-molecule fluorescence probes dynamics of barrier crossing. *Nature* **2013**, *502*, 685−688.
(53) Kim, W. K.; Netz, R. R. The mean shape of transition and first-passage paths. *J. Chem. Phys.* **2015**, *143* (22), 224108.
(54) Hummer, G. From transition paths to transition states and rate coefficients. *J. Chem. Phys.* **2004**, *120* (2), 516−523.
(55) Sturzenegger, F.; Zosel, F.; Holmstrom, E. D.; Buholzer, K. J.; Makarov, D. E.; Nettels, D.; Schuler, B. Transition path times of coupled folding and binding reveal the formation of an encounter complex. *Nat. Commun.* **2018**, *9* (1), 4708.
(56) Kim, J.-Y.; Chung, H. S. Disordered proteins follow diverse transition paths as they fold and bind to a partner. *Science* **2020**, *368* (6496), 1253−1257.
(57) Ramanathan, R.; Muñoz, V. A Method for Extracting the Free Energy Surface and Conformational Dynamics of Fast-Folding Proteins from Single Molecule Photon Trajectories. *J. Phys. Chem. B* **2015**, *119* (25), 7944−56.
(58) Taumoefolau, G. H.; Best, R. B. Estimating transition path times and shapes from single-molecule photon trajectories: A simulation analysis. *J. Chem. Phys.* **2021**, *154* (11), 115101.
(59) Muñoz, V.; Sanchez-Ruiz, J. M. Exploring protein-folding ensembles: a variable-barrier model for the analysis of equilibrium unfolding experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (51), 17646−51.
(60) Fung, A.; Li, P.; Godoy-Ruiz, R.; Sanchez-Ruiz, J. M.; Muñoz, V. Expanding the realm of ultrafast protein folding: gpW, a midsize natural single-domain with alpha+beta topology that folds downhill. *J. Am. Chem. Soc.* **2008**, *130* (23), 7489−95.
(61) Schuler, B.; Lipman, E. A.; Eaton, W. A. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **2002**, *419* (6908), 743−747.
(62) Campos, L. A.; Sadqi, M.; Liu, J. W.; Wang, X.; English, D. S.; Muñoz, V. Gradual Disordering of the Native State on a Slow Two-State Folding Protein Monitored by Single-Molecule Fluorescence Spectroscopy and NMR. *J. Phys. Chem. B* **2013**, *117* (42), 13120−13131.
(63) Chung, H. S.; Gopich, I. V. Fast single-molecule FRET spectroscopy: theory and experiment. *Phys. Chem. Chem. Phys.* **2014**, *16* (35), 18644−18657.
(64) Gopich, I. V.; Szabo, A. Single-molecule FRET with diffusion and conformational dynamics. *J. Phys. Chem. B* **2007**, *111* (44), 12925−12932.
(65) Muñoz, V. Conformational dynamics and ensembles in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 395−412.
(66) Naganathan, A. N.; Muñoz, V. Insights into protein folding mechanisms from large scale analysis of mutational effects. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (19), 8611−6.
(67) Jackson, S. E. How do small single-domain proteins fold? *Folding Des.* **1998**, *3* (4), R81−R91.
(68) Maxwell, K. L.; Wildes, D.; Zarrine-Afsar, A.; De Los Rios, M. A.; Brown, A. G.; Friel, C. T.; Hedberg, L.; Horng, J. C.; Bona, D.; Miller, E. J.; et al. Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci.* **2005**, *14* (3), 602−616.
(69) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S. On the transition coordinate for protein folding. *J. Chem. Phys.* **1998**, *108* (1), 334−350.
(70) Campos, L. A.; Liu, J.; Wang, X.; Ramanathan, R.; English, D. S.; Muñoz, V. A photoprotection strategy for microsecond-resolution single-molecule fluorescence spectroscopy. *Nat. Methods* **2011**, *8* (2), 143−6.
(71) Hummer, G. Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New J. Phys.* **2005**, *7* (1), 34.

https://doi.org/10.1021/acs.jpcb.1c05401
J. Phys. Chem. B 2021, 125, 12413−12425