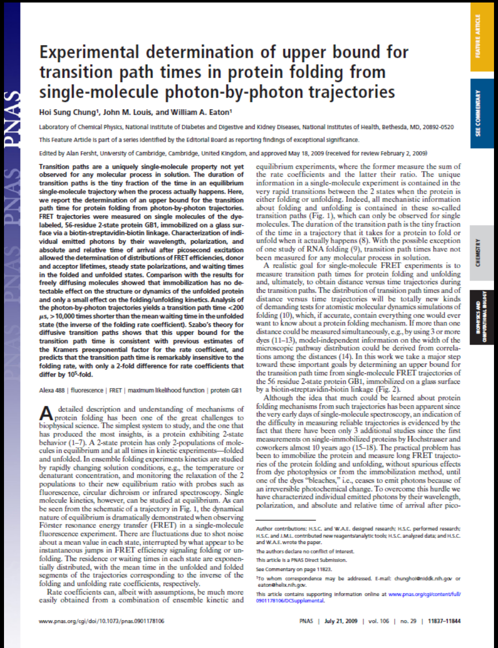
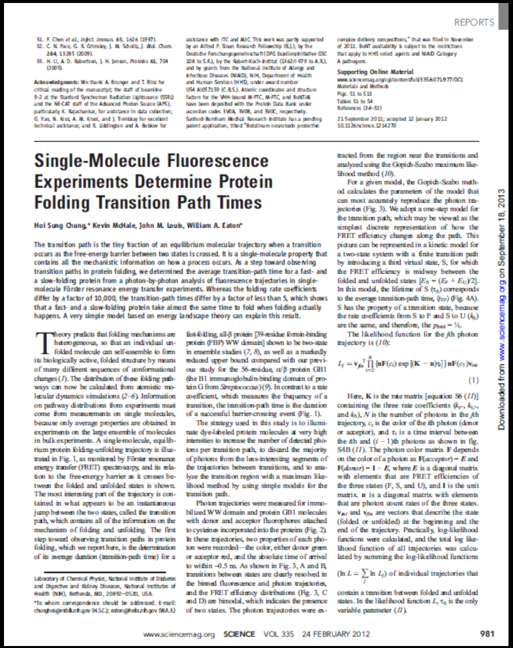


Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories

Hoi Sung Chung[1], John M. Louis, and William A. Eaton[1]

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, 20892-0520

www.pnas.org/cgi/doi/10.1073/pnas.0901178106 PNAS | July 21, 2009 | vol. 106 | no. 29 | 11837–11844



Single-Molecule Fluorescence Experiments Determine Protein Folding Transition Path Times

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

Chemical Physics 422 (2013) 229–237

Contents lists available at SciVerse ScienceDirect

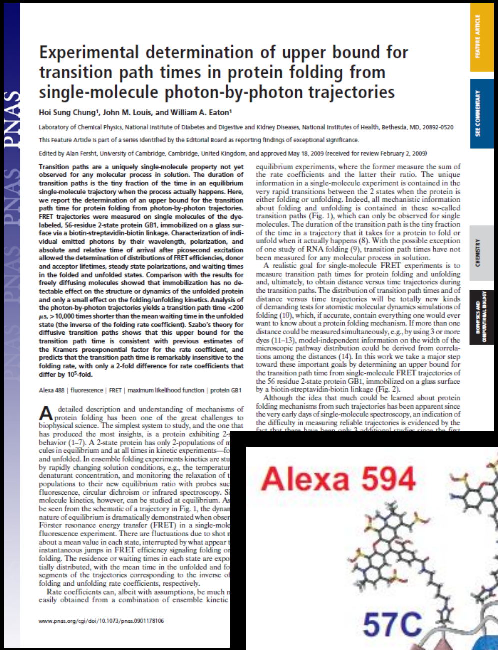
Chemical Physics

journal homepage: www.elsevier.com/locate/chemphys

Measuring ultrafast protein folding rates from photon-by-photon analysis of single molecule fluorescence trajectories

Hoi Sung Chung *, Troy Cellmer, John M. Louis, William A. Eaton *

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520, USA



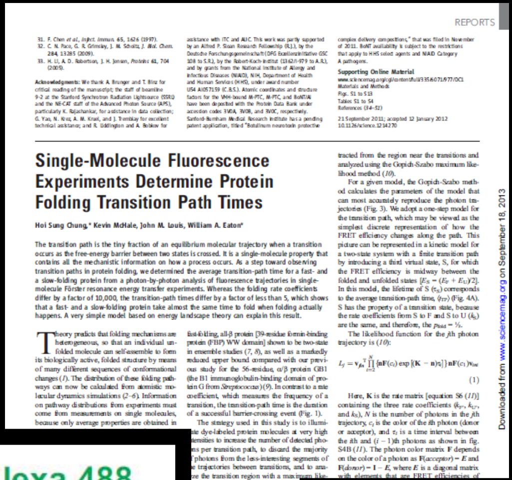
PNAS

Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories

Hoi Sung Chung[1], John M. Louis, and William A. Eaton[1]

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, 20892-0520

www.pnas.org/cgi/doi/10.1073/pnas.0901178106



REPORTS

Single-Molecule Fluorescence Experiments Determine Protein Folding Transition Path Times

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

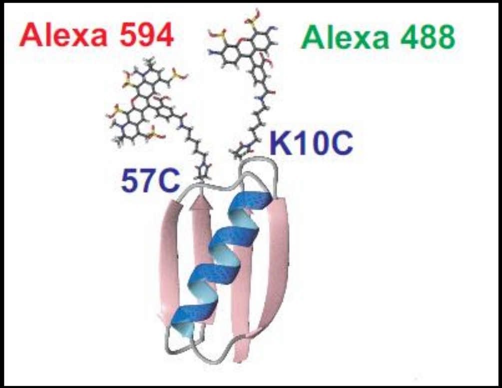
Chemical Physics 422 (2013) 229–237

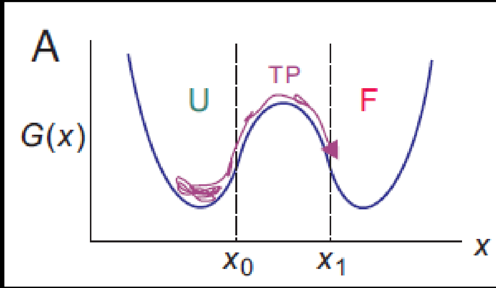
Chemical Physics

journal homepage: www.elsevier.com/locate/chemphys

Measuring ultrafast protein folding rates from photon-by-photon analysis of single molecule fluorescence trajectories

Hoi Sung Chung *, Troy Cellmer, John M. Louis, William A. Eaton *





PNAS PNAS PNAS

Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories

Hoi Sung Chung[1], John M. Louis, and William A. Eaton[1]

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, 20892-0520

This Feature Article is part of a series identified by the Editorial Board as reporting findings of exceptional significance.

Edited by Alan Fersht, University of Cambridge, Cambridge, United Kingdom, and approved May 18, 2009 (received for review February 2, 2009)

Transition paths are a uniquely single-molecule property not yet observed for any molecular process in solution. The duration of transition paths is the tiny fraction of the time in an equilibrium single-molecule trajectory when the process actually happens. Here, we report the determination of an upper bound for the transition path time for protein folding from photon-by-photon trajectories. FRET trajectories were measured on single molecules of the dye-labeled, 56-residue 2-state protein GB1, immobilized on a glass surface via a biotin-streptavidin-biotin linkage. Characterization of individual emitted photons by their wavelength, polarization, and absolute and relative time of arrival after picosecond excitation allowed the determination of distributions of FRET efficiencies, donor and acceptor lifetimes, steady state polarizations, and waiting times in the folded and unfolded states. Comparison with the results for freely diffusing molecules showed that immobilization has no detectable effect on the structure or dynamics of the unfolded protein and only a small effect on the folding/unfolding kinetics. Analysis of the photon-by-photon trajectories yields a transition path time <200 μs, >10,000 times shorter than the mean waiting time in the unfolded state (the inverse of the folding rate coefficient). Szabo's theory for diffusive transition paths shows that this upper bound for the transition path time is consistent with previous estimates of the Kramers preexponential factor for the rate coefficient, and predicts that the transition path time is remarkably insensitive to the folding rate, with only a 2-fold difference for rate coefficients that differ by 10^5-fold.

Alexa 488 | fluorescence | FRET | maximum likelihood function | protein GB1

A detailed description and understanding of mechanisms of protein folding has been one of the great challenges to biophysical science. The simplest system to study, and the one that has produced the most insights, is a protein exhibiting 2-state behavior (1–7). A 2-state protein has only 2-populations of molecules in equilibrium and at all times in kinetic experiments—folded and unfolded. In ensemble folding experiments kinetics are studied by rapidly changing solution conditions, e.g., the temperature or denaturant concentration, and monitoring the relaxation of the 2 populations to their new equilibrium ratio with probes such as fluorescence, circular dichroism or infrared spectroscopy. Single molecule kinetics, however, can be studied at equilibrium. As can be seen from the schematic of a trajectory in Fig. 1, the dynamical nature of equilibrium is dramatically demonstrated when observing Förster resonance energy transfer (FRET) in a single-molecule fluorescence experiment. There are fluctuations due to shot noise about a mean value in each state, interrupted by what appear to be instantaneous jumps in FRET efficiency signaling folding or unfolding. The residence or waiting times in each state are exponentially distributed, with the mean time in the unfolded and folded segments of the trajectories corresponding to the inverse of the folding and unfolding rate coefficients, respectively.

Rate coefficients can, albeit with assumptions, be much more easily obtained from a combination of ensemble kinetic and equilibrium experiments, where the former measure the sum of the rate coefficients and the latter their ratio. The unique information in a single-molecule experiment is contained in the very rapid transitions between the 2 states when the protein is either folding or unfolding. Indeed, all mechanistic information about folding and unfolding is contained in these so-called transition paths (Fig. 1), which can only be observed for single molecules. The duration of the transition path is the tiny fraction of the time in a trajectory that it takes for a protein to fold or unfold when it actually happens (8). With the possible exception of one study of RNA folding (9), transition path times have not been measured for any molecular process in solution.

A realistic goal for single-molecule FRET experiments is to measure transition path times for protein folding and unfolding and, ultimately, to obtain distance versus time trajectories during the transition paths. The distribution of transition path times and of distance versus time trajectories will be totally new kinds of demanding tests for atomistic molecular dynamics simulations of folding (10), which, if accurate, contain everything one would ever want to know about a protein folding mechanism. If more than one distance could be measured simultaneously, e.g., by using 3 or more dyes (11–13), model-independent information on the width of the microscopic pathway distribution could be derived from correlations among the distances (14). In this work we take a major step toward these important goals by determining an upper bound for the transition path time from single-molecule FRET trajectories of the 56 residue 2-state protein GB1, immobilized on a glass surface by a biotin-streptavidin-biotin linkage (Fig. 2).

Although the idea that much could be learned about protein folding mechanisms from such trajectories has been apparent since the very early days of single-molecule spectroscopy, an indication of the difficulty in measuring reliable trajectories is evidenced by the fact that there have been only 3 additional studies since the first measurements on single-immobilized proteins by Hochstrasser and coworkers almost 10 years ago (15–18). The practical problem has been to immobilize the protein and measure long FRET trajectories of the protein folding and unfolding, without spurious effects from dye photophysics or from the immobilization method, until one of the dyes "bleaches," i.e., ceases to emit photons because of an irreversible photochemical change. To overcome this hurdle we have characterized individual emitted photons by their wavelength, polarization, and absolute and relative time of arrival after pico-

Author contributions: H.S.C. and W.A.E. designed research; H.S.C. performed research; H.S.C. and J.M.L. contributed new reagents/analytic tools; H.S.C. analyzed data; and H.S.C. and W.A.E. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

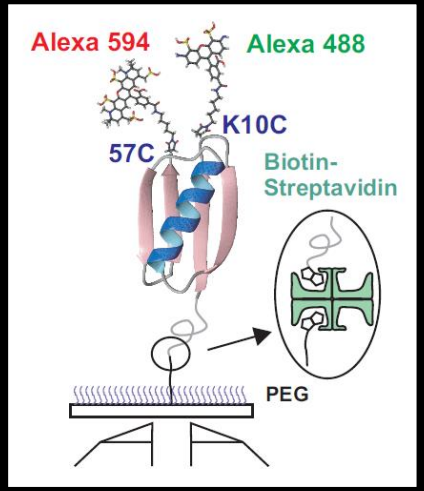
See Commentary on page 11823.

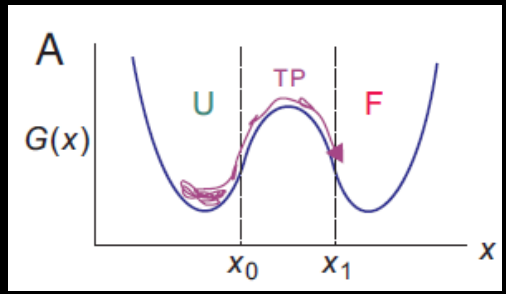
[1]To whom correspondence may be addressed. E-mail: chunghoi@niddk.nih.gov or eaton@helix.nih.gov.

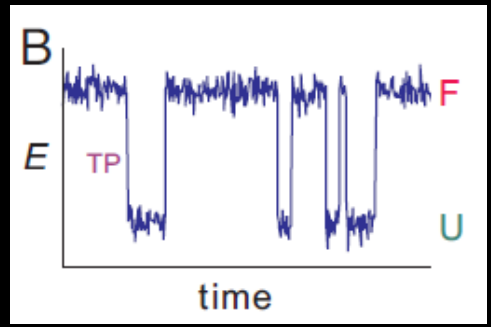
This article contains supporting information online at www.pnas.org/cgi/content/full/0901178106/DCSupplemental.

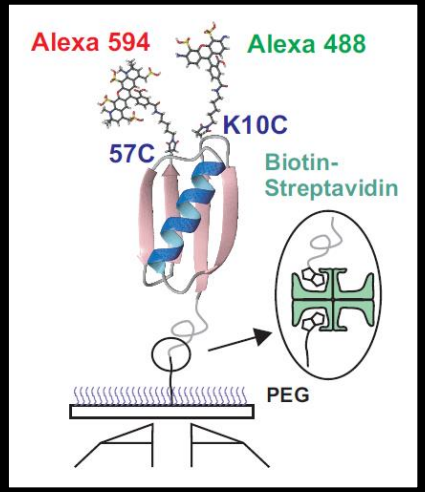
Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories

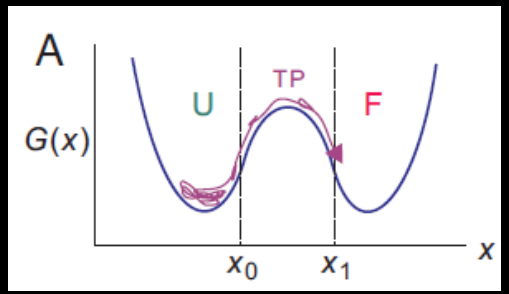
Hoi Sung Chung[1], John M. Louis, and William A. Eaton[1]

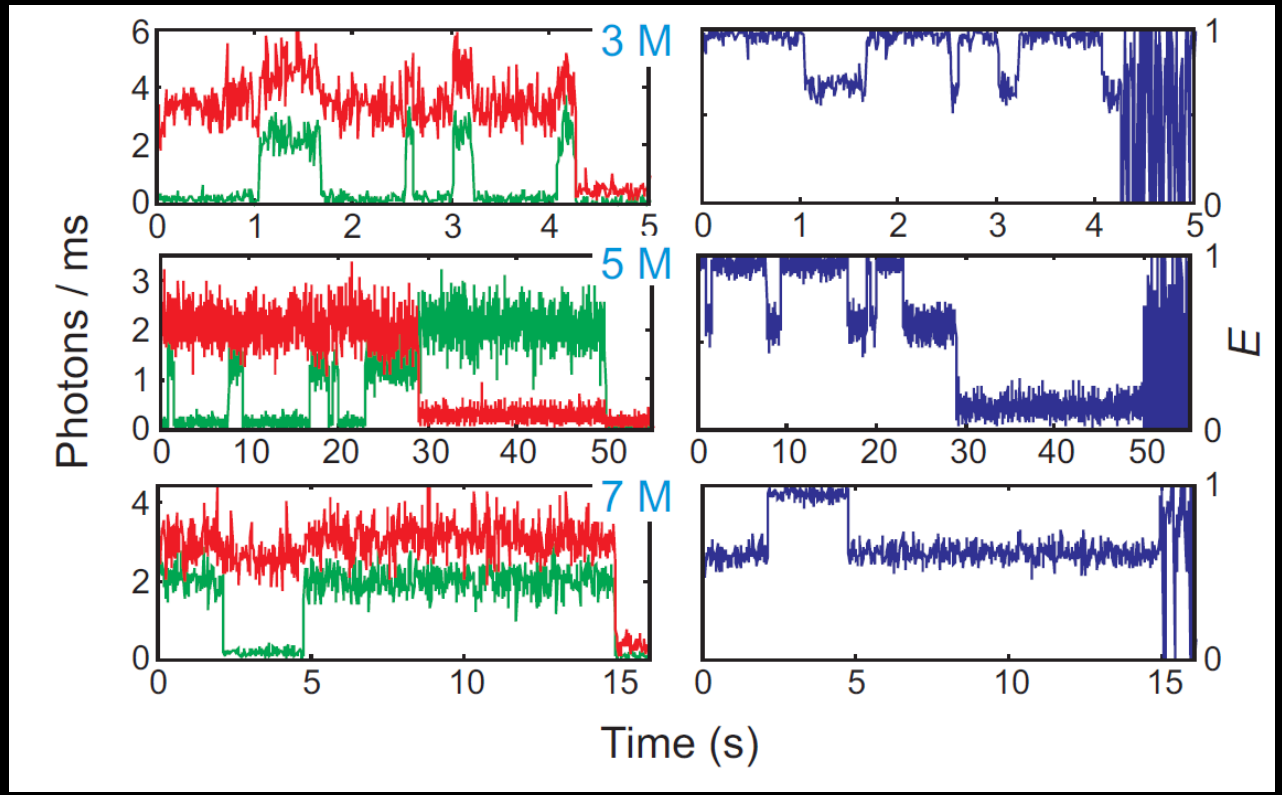


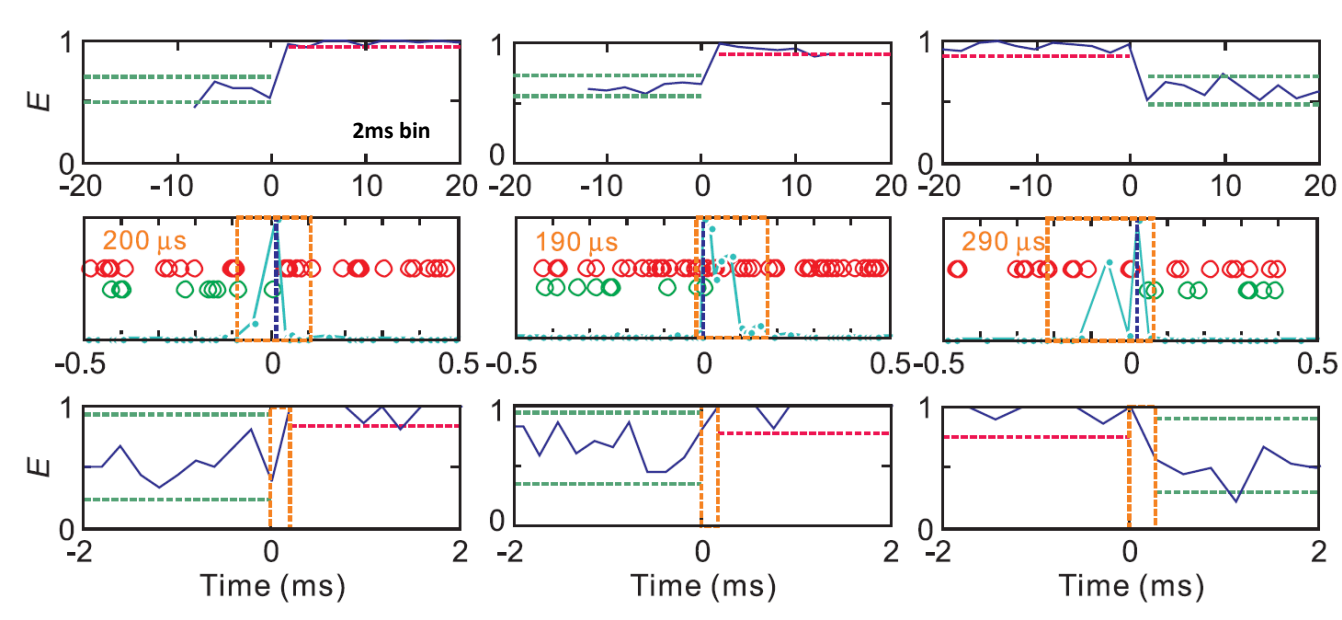





Alexa 594
Alexa 488
K10C
57C
Biotin-
Streptavidin
PEG


A
G(x)
U
TP
F
x_0
x_1
x


3 M
5 M
7 M
Photons / ms
E
Time (s)



$f_{F(U)}(n)$: the probability of n photons belonging to the folded (unfolded) state

$$f_{F(U)}(n) = E_{F(U)}^{n_A}\left(1 - E_{F(U)}\right)^{n-n_A}\prod_{j=1}^{n-1} P_{F(U)}(d_j)$$

$$P_{F(U)}(d) = \nu_{F(U)} e^{-\nu_{F(U)} d}$$

d : the time interval

$\nu_{F(U)}$: photon counting rates

31. F. Chen *et al.*, *Infect. Immun.* 65, 1626 (1997).
32. C. N. Pace, G. R. Grimsley, J. M. Scholtz, *J. Biol. Chem.* 284, 13285 (2009).
33. H. Li, A. D. Robertson, J. H. Jensen, *Proteins* 61, 704 (2005).

Acknowledgments: We thank A. Brunger and T. Binz for critical reading of the manuscript; the staff of beamline 9-2 at the Stanford Synchrotron Radiation Lightsource (SSRL) and the NE-CAT staff of the Advanced Photon Source (APS), particularly K. Rajashankar, for assistance in data collection; G. Yao, N. Kreg, A. M. Kruel, and J. Tremblay for excellent technical assistance; and R. Liddington and A. Bobkov for assistance with ITC and AUC. This work was partly supported by an Alfred P. Sloan Research Fellowship (R.J.), by the Deutsche Forschungsgemeinschaft (DFG Exzellenzinitiative GSC 108 to S.R.), by the Robert-Koch-Institut (1362/I-979 to A.R.), and by grants from the National Institute of Allergy and Infectious Diseases (NIAID), NIH, Department of Health and Human Services (HHS), under award number U54 AI057159 (C.B.S.). Atomic coordinates and structure factors for the VHH-bound M-PTC, M-PTC, and BoNT/AI have been deposited with the Protein Data Bank under accession codes 3V0A, 3V0B, and 3V0C, respectively. Sanford-Burnham Medical Research Institute has a pending patent application, titled "Botulinum neurotoxin protective complex delivery compositions," that was filed in November of 2011. BoNT availability is subject to the restrictions that apply to HHS select agents and NIAID Category A pathogens.

Supporting Online Material
www.sciencemag.org/cgi/content/full/335/6071/977/DC1
Materials and Methods
Figs. S1 to S13
Tables S1 to S4
References (34–51)

21 September 2011; accepted 12 January 2012
10.1126/science.1214270

Single-Molecule Fluorescence Experiments Determine Protein Folding Transition Path Times

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

The transition path is the tiny fraction of an equilibrium molecular trajectory when a transition occurs as the free-energy barrier between two states is crossed. It is a single-molecule property that contains all the mechanistic information on how a process occurs. As a step toward observing transition paths in protein folding, we determined the average transition-path time for a fast- and a slow-folding protein from a photon-by-photon analysis of fluorescence trajectories in single-molecule Förster resonance energy transfer experiments. Whereas the folding rate coefficients differ by a factor of 10,000, the transition-path times differ by a factor of less than 5, which shows that a fast- and a slow-folding protein take almost the same time to fold when folding actually happens. A very simple model based on energy landscape theory can explain this result.

Theory predicts that folding mechanisms are heterogeneous, so that an individual unfolded molecule can self-assemble to form its biologically active, folded structure by means of many different sequences of conformational changes (*1*). The distribution of these folding pathways can now be calculated from atomistic molecular dynamics simulations (*2*–*6*). Information on pathway distributions from experiments must come from measurements on single molecules, because only average properties are obtained in experiments on the large ensemble of molecules in bulk experiments. A single-molecule, equilibrium protein folding-unfolding trajectory is illustrated in Fig. 1, as monitored by Förster resonance energy transfer (FRET) spectroscopy, and its relation to the free-energy barrier as it crosses between the folded and unfolded states is shown. The most interesting part of the trajectory is contained in what appears to be an instantaneous jump between the two states, called the transition path, which contains all of the information on the mechanism of folding and unfolding. The first step toward observing transition paths in protein folding, which we report here, is the determination of its average duration (transition-path time) for a fast-folding, all-β protein [39-residue formin-binding protein (FBP) WW domain] shown to be two-state in ensemble studies (*7*, *8*), as well as a markedly reduced upper bound compared with our previous study for the 56-residue, α/β protein GB1 (the B1 immunoglobulin-binding domain of protein G from *Streptococcus*) (*9*). In contrast to a rate coefficient, which measures the frequency of a transition, the transition-path time is the duration of a successful barrier-crossing event (Fig. 1).

The strategy used in this study is to illuminate dye-labeled protein molecules at very high intensities to increase the number of detected photons per transition path, to discard the majority of photons from the less-interesting segments of the trajectories between transitions, and to analyze the transition region with a maximum likelihood method by using simple models for the transition path.

Photon trajectories were measured for immobilized WW domain and protein GB1 molecules with donor and acceptor fluorophores attached to cysteines incorporated into the proteins (Fig. 2). In these trajectories, two properties of each photon were recorded—the color, either donor green or acceptor red, and the absolute time of arrival to within ~0.5 ns. As shown in Fig. 3, A and B, transitions between states are clearly resolved in the binned fluorescence and photon trajectories, and the FRET efficiency distributions (Fig. 3, C and D) are bimodal, which indicates the presence of two states. The photon trajectories were extracted from the region near the transitions and analyzed using the Gopich-Szabo maximum likelihood method (*10*).

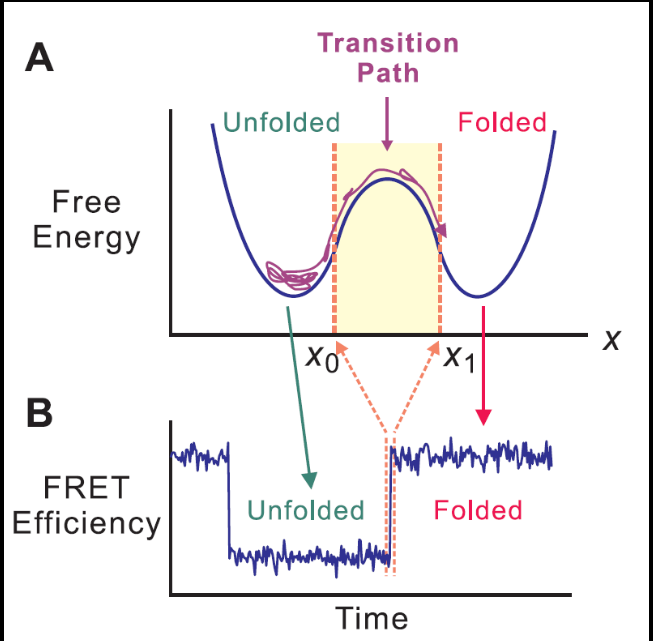
For a given model, the Gopich-Szabo method calculates the parameters of the model that can most accurately reproduce the photon trajectories (Fig. 3). We adopt a one-step model for the transition path, which may be viewed as the simplest discrete representation of how the FRET efficiency changes along the path. This picture can be represented in a kinetic model for a two-state system with a finite transition path by introducing a third virtual state, S, for which the FRET efficiency is midway between the folded and unfolded states [$E_S = (E_F + E_U)/2$]. In this model, the lifetime of S (τ_S) corresponds to the average transition-path time, $\langle t_{TP} \rangle$ (Fig. 4A). S has the property of a transition state, because the rate coefficients from S to F and S to U (k_S) are the same, and therefore, the $p_{fold} = ½$.

The likelihood function for the *j*th photon trajectory is (*10*):

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

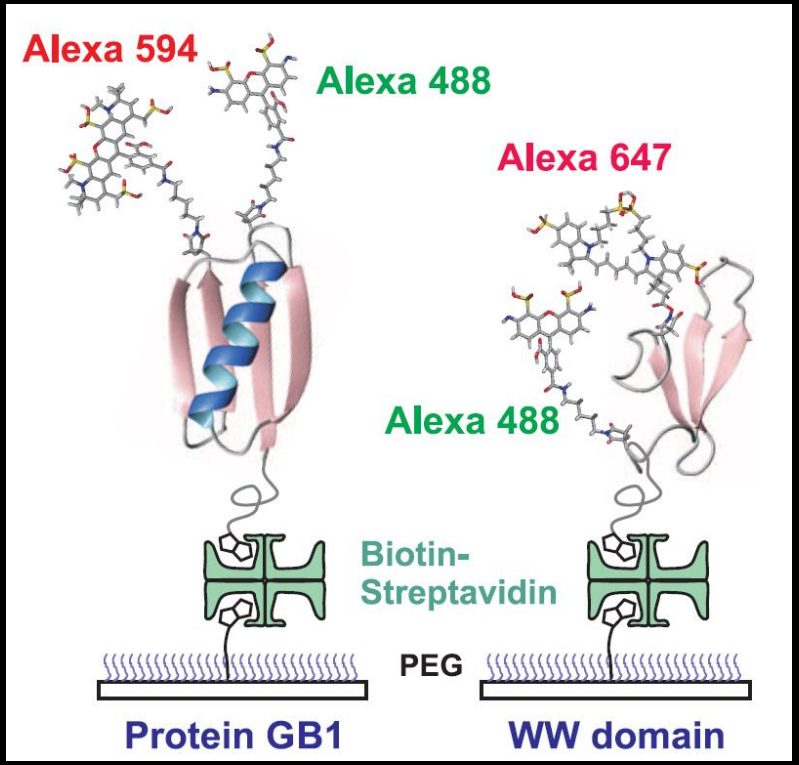
Here, **K** is the rate matrix [equation S6 (*11*)] containing the three rate coefficients (k_F, k_U, and k_S), N is the number of photons in the *j*th trajectory, c_i is the color of the *i*th photon (donor or acceptor), and τ_i is a time interval between the *i*th and $(i-1)$th photons as shown in fig. S4B (*11*). The photon color matrix **F** depends on the color of a photon as $\mathbf{F}(acceptor) = E$ and $\mathbf{F}(donor) = \mathbf{I} - E$, where E is a diagonal matrix with elements that are FRET efficiencies of the three states (F, S, and U), and **I** is the unit matrix. **n** is a diagonal matrix with elements that are photon count rates of the three states. $\mathbf{v}_{ini}$ and $\mathbf{v}_{fin}$ are vectors that describe the state (folded or unfolded) at the beginning and the end of the trajectory. Practically, log-likelihood functions were calculated, and the total log likelihood function of all trajectories was calculated by summing the log-likelihood functions ($\ln L = \sum_j \ln L_j$) of individual trajectories that contain a transition between folded and unfolded states. In the likelihood function L, τ_S is the only variable parameter (*11*).

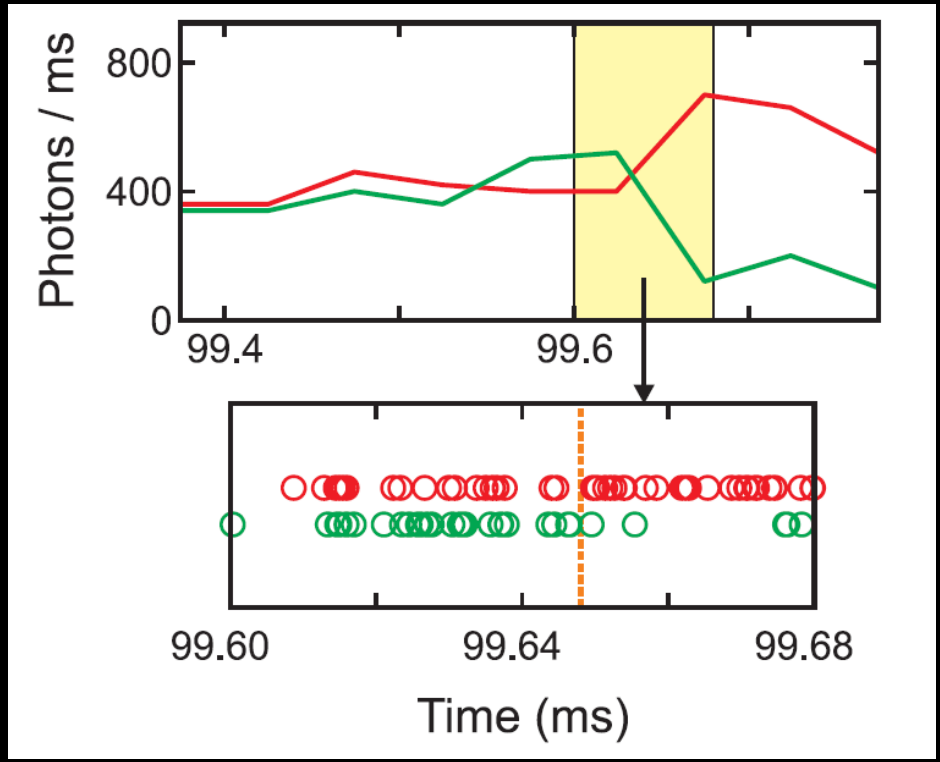
Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health (NIH), Bethesda, MD, 20892–0520, USA.
*To whom correspondence should be addressed. E-mail: chunghoi@niddk.nih.gov (H.S.C.); eaton@helix.nih.gov (W.A.E.)

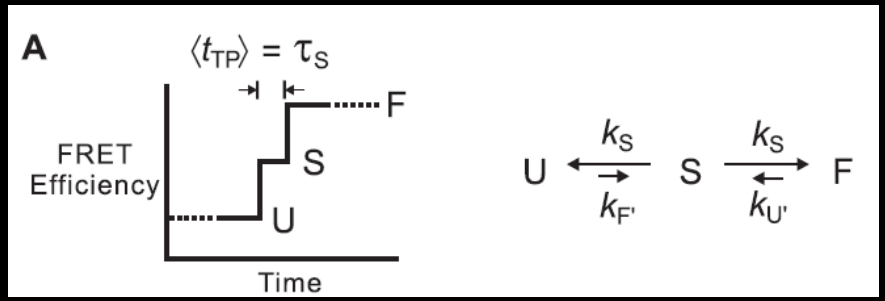


Single-Molecule Fluorescence Experiments Determine Protein Folding Transition Path Times

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

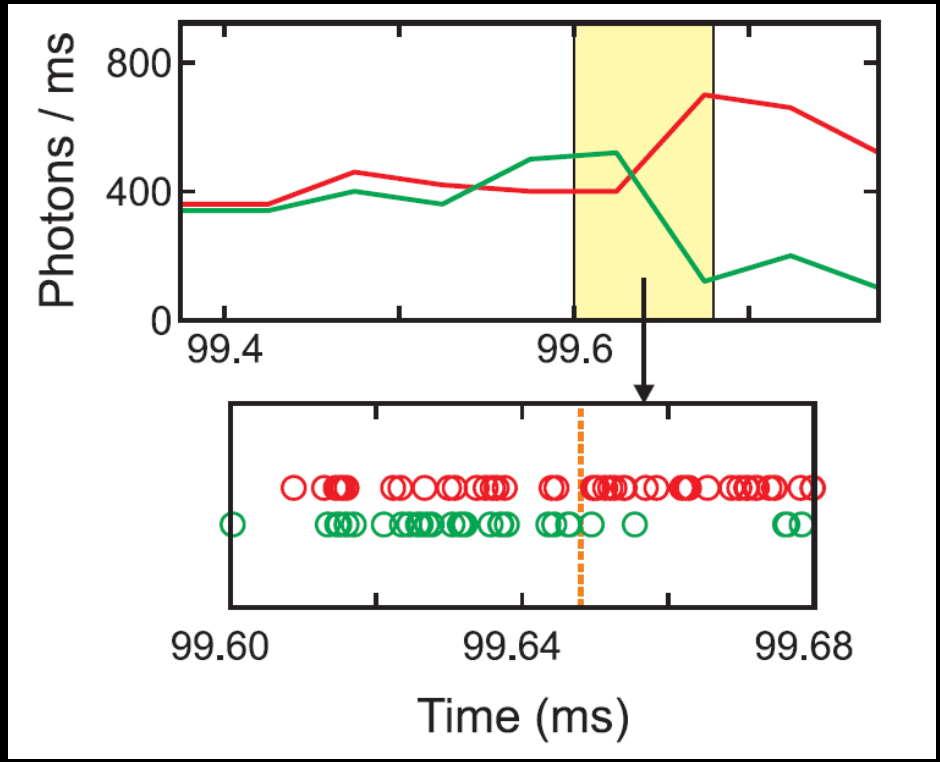

Alexa 594
Alexa 488
Alexa 647
Alexa 488
Biotin-
Streptavidin
PEG
Protein GB1
WW domain

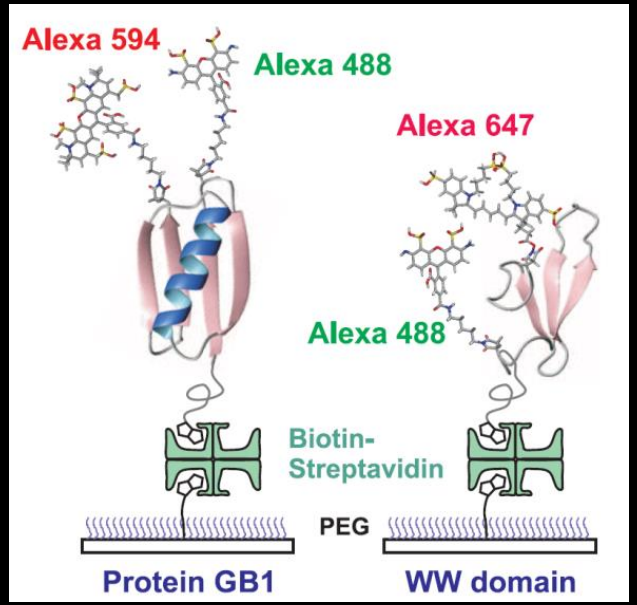

Photons / ms
800
400
0
99.4
99.6
99.60
99.64
99.68
Time (ms)

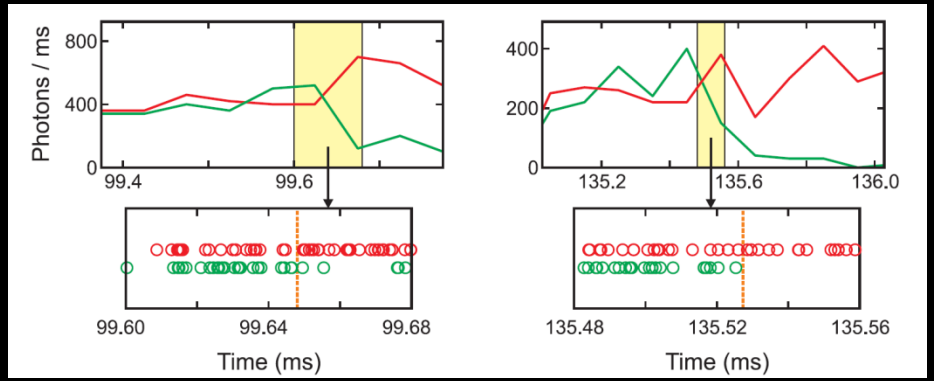


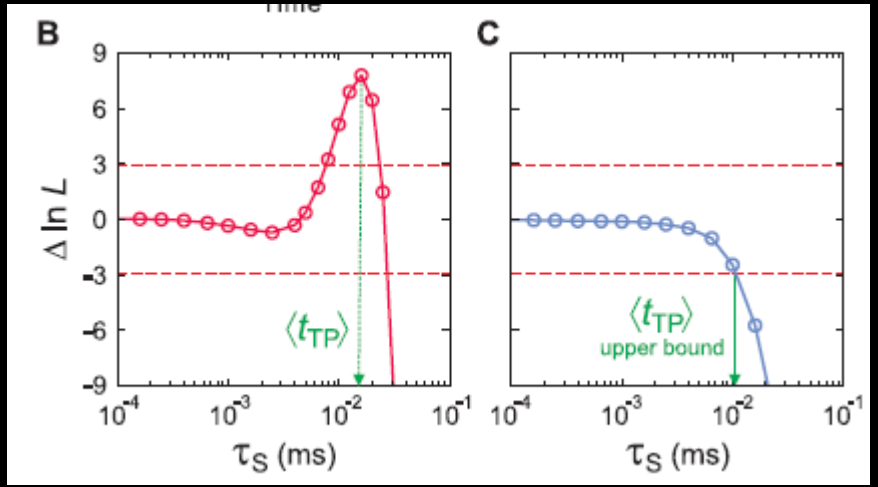
$$L_j = \mathbf{1}^T \prod_{i=2}^{N_j} [\mathbf{F}(c_i) \exp(\mathbf{K}\tau_i)] \mathbf{F}(c_1) \mathbf{p}_{eq},$$

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




Alexa 594
Alexa 488
Alexa 647
Alexa 488
Biotin-
Streptavidin
PEG
Protein GB1
WW domain


Photons / ms
800
400
0
99.4
99.6
99.60
99.64
99.68
Time (ms)
400
200
0
135.2
135.6
136.0
135.48
135.52
135.56
Time (ms)


B
C
Δ ln L
9
6
3
0
-3
-6
-9
⟨tTP⟩
τS (ms)
⟨tTP⟩
upper bound
τS (ms)

Chemical Physics 422 (2013) 229–237

Contents lists available at SciVerse ScienceDirect

Chemical Physics

journal homepage: www.elsevier.com/locate/chemphys

Measuring ultrafast protein folding rates from photon-by-photon analysis of single molecule fluorescence trajectories

Hoi Sung Chung *, Troy Cellmer, John M. Louis, William A. Eaton *

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520, USA

ARTICLE INFO

Article history:
Available online 14 August 2012

Keywords:
Single-molecule fluorescence
FRET
Maximum likelihood analysis
Dye blinking
Ultrafast protein folding rates
Photon trajectories
Villin subdomain

ABSTRACT

Folding and unfolding rates for the ultrafast folding villin subdomain were determined from a photon-by-photon analysis of fluorescence trajectories in single molecule FRET experiments. One of the obstacles to measuring fast kinetics in single molecule fluorescence experiments is blinking of the fluorophores on a timescale that is not well separated from the process of interest. By incorporating acceptor blinking into a two-state kinetics model, we show that it is possible to extract accurate rate coefficients on the micro-second time scale for folding and unfolding using the maximum likelihood method of Gopich and Szabo. This method yields the most likely parameters of a given model that can reproduce the observed photon trajectories. The extracted parameters agree with both the decay rate of the donor-acceptor cross correlation function and the results of ensemble equilibrium and kinetic experiments using nanosecond laser temperature jump.

Published by Elsevier B.V.

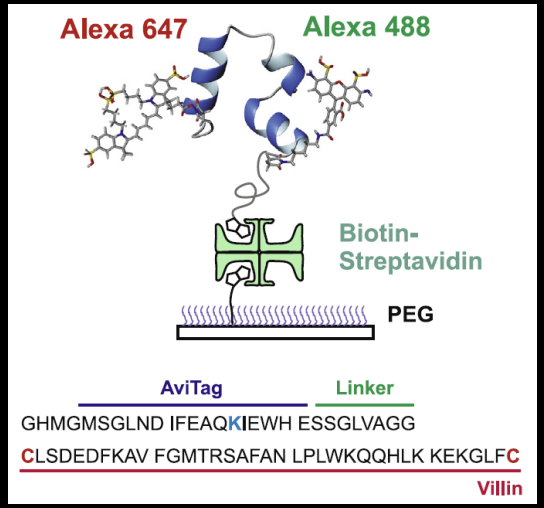
1. Introduction

The power of single molecule fluorescence spectroscopy is the ability to investigate distributions in molecular behavior for intrinsically heterogeneous systems. One such system is protein folding, in which theory predicts many different sequences of structural changes in the pathways that connect the folded and unfolded states [1]. The α-helical, 35-residue villin subdomain (Fig. 1) is currently the most extensively studied protein by experiment, theory, and simulations (see bibliography in Supplementary Material). The reasons are that it has equilibrium properties of a much larger single-domain protein [2,3], is among the fastest folding proteins [4,5], and exhibits unusual kinetics such as a denaturant independent relaxation rate [6] and an apparent increase in the internal friction with temperature in a Kramers description of the barrier crossing [7]. Our ultimate goal for single molecule experiments on this protein is to observe the distribution of transition paths – a uniquely single molecule property. Such measurements would provide a very demanding test of the accuracy of the mechanisms found in molecular dynamics simulations, but represent a major challenge since transition paths have not been observed for any molecular system in the condensed phase. New and sensitive tests of simulation are important because, if accurate, everything one would ever want to know about the folding mechanism of a particular protein is contained in a sufficiently long atomistic trajectory [8–10] or Markov state modeling of many short trajectories [11].

Studies of the villin subdomain are particularly challenging because of its very rapid kinetics, with folding times of the wild-type on the order of 5–50 μs [3,6,12]. Until quite recently the time resolution in single molecule FRET experiments [13] has been limited by the bin time of the measurement, which is usually 1–10 ms at the moderate illumination intensities that have been employed to avoid photochemical problems such as bleaching and blinking of the dyes [14]. For residence times much longer than these bin times, two distinct peaks will appear in a histogram of the FRET

* Corresponding authors.
E-mail addresses: chungh@niddk.nih.gov (H.S. Chung), eaton@helix.nih.gov (W.A. Eaton).

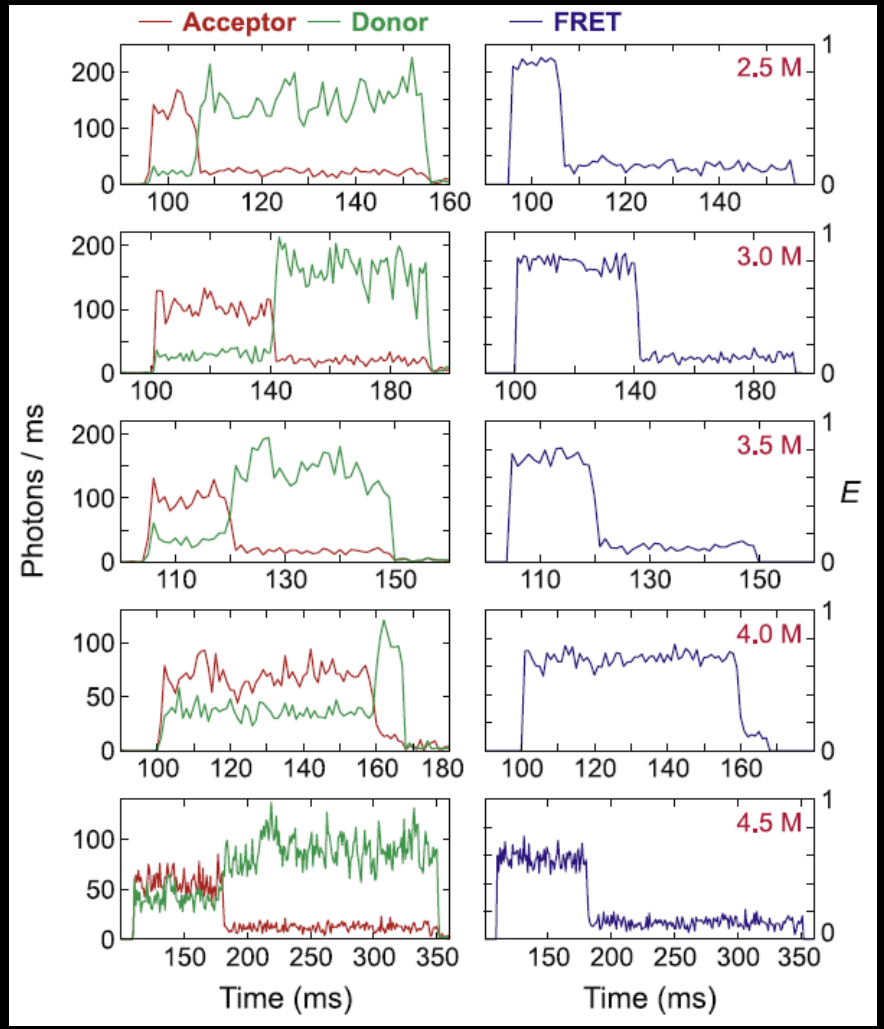
0301-0104/$ - see front matter Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.chemphys.2012.08.005



Measuring ultrafast protein folding rates from photon-by-photon analysis of single molecule fluorescence trajectories

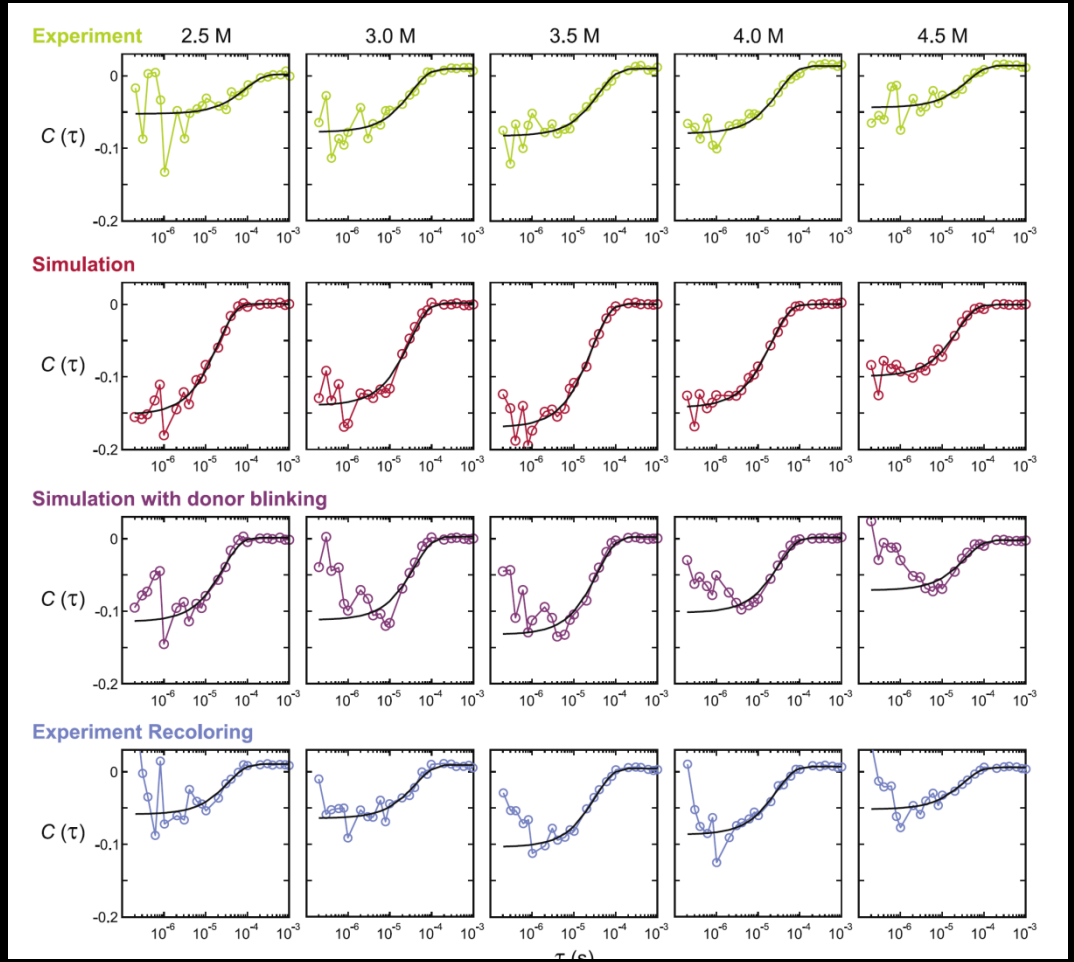
Hoi Sung Chung *, Troy Cellmer, John M. Louis, William A. Eaton *

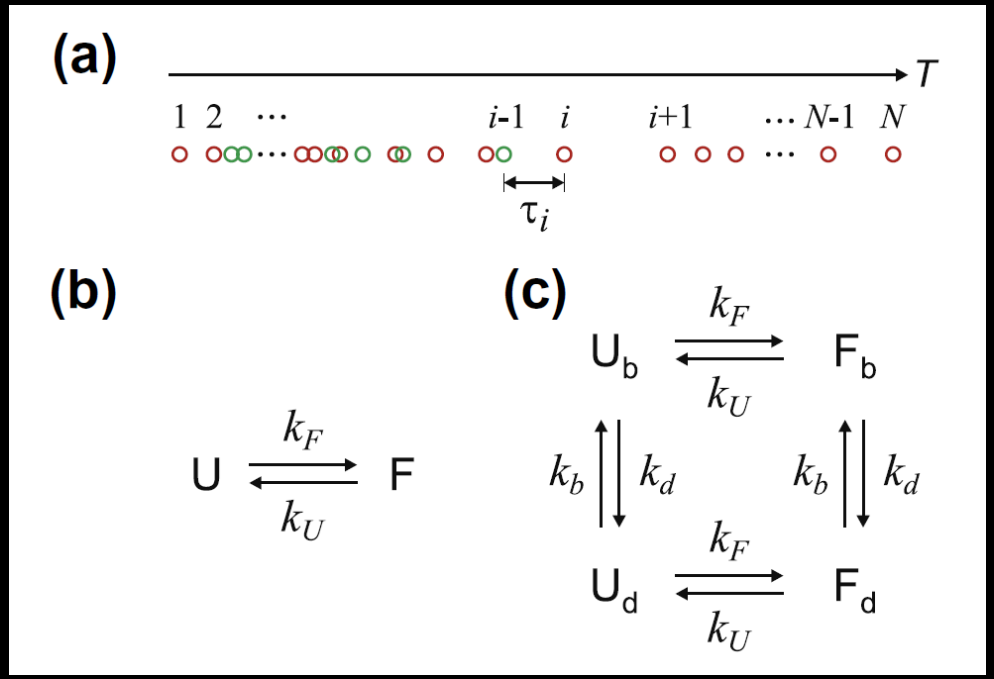
Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520, USA



$$L_j = \mathbf{1}^T \prod_{i=2}^{N_j} [\mathbf{F}(c_i) \exp(\mathbf{K}\tau_i)] \mathbf{F}(c_1) \mathbf{p}_{eq},$$

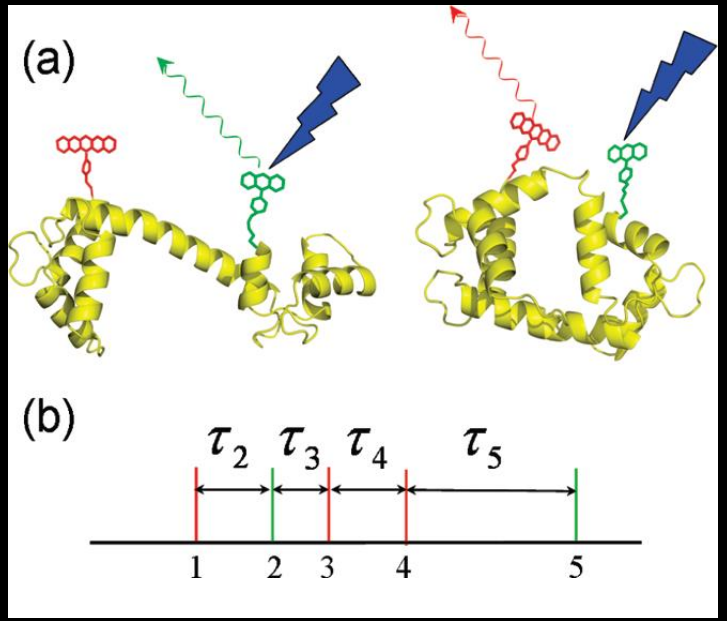
$$\mathbf{K} = \begin{pmatrix} -k_U & k_F \\ k_U & -k_F \end{pmatrix}$$


Experiment
2.5 M
3.0 M
3.5 M
4.0 M
4.5 M
Simulation
Simulation with donor blinking
Experiment Recoloring
C (τ)
0
-0.1
-0.2
10-6
10-5
10-4
10-3



$$\mathbf{K} = \begin{pmatrix} -k_U & k_F \\ k_U & -k_F \end{pmatrix}$$

$$\mathbf{K} = \begin{pmatrix} -k_U - k_d & k_F & k_b & 0 \\ k_U & -k_F - k_d & 0 & k_b \\ k_d & 0 & -k_U - k_b & k_F \\ 0 & k_d & k_U & -k_F - k_b \end{pmatrix}$$



$$L = \mathbf{1}^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{E})\mathrm{e}^{\boldsymbol{K}\tau_5}\boldsymbol{E}\mathrm{e}^{\boldsymbol{K}\tau_4}\boldsymbol{E}\mathrm{e}^{\boldsymbol{K}\tau_3}(\boldsymbol{I} - \boldsymbol{E})\mathrm{e}^{\boldsymbol{K}\tau_2}\boldsymbol{E}\boldsymbol{p}_{\mathrm{eq}}$$

For a two-state case & two detected photon,

$$L = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1-E_1 & 0 \\ 0 & 1-E_2 \end{bmatrix} \begin{bmatrix} e^{-k_{21}\cdot\tau_2} & e^{k_{12}\cdot\tau_2} \\ e^{k_{21}\cdot\tau_2} & e^{-k_{12}\cdot\tau_2} \end{bmatrix} \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$$

$$L = (1-E_1)e^{-k_{21}\cdot\tau_2}E_1P_1 + (1-E_1)e^{k_{12}\cdot\tau_2}E_2P_2$$

$$+ (1-E_2)e^{k_{21}\cdot\tau_2}E_1P_1 + (1-E_2)e^{-k_{12}\cdot\tau_2}E_2P_2$$

$$L = \mathbf{1}^{\mathrm{T}} \prod_{k=2}^{N_{\mathrm{ph}}} (\boldsymbol{F}(c_k)\mathrm{e}^{\boldsymbol{K}\tau_k})\boldsymbol{F}(c_1)\boldsymbol{p}_{\mathrm{eq}}$$