National Institutes of…..

Advantages (+) and disadvantages (-) of working with the nine types of principal investigators

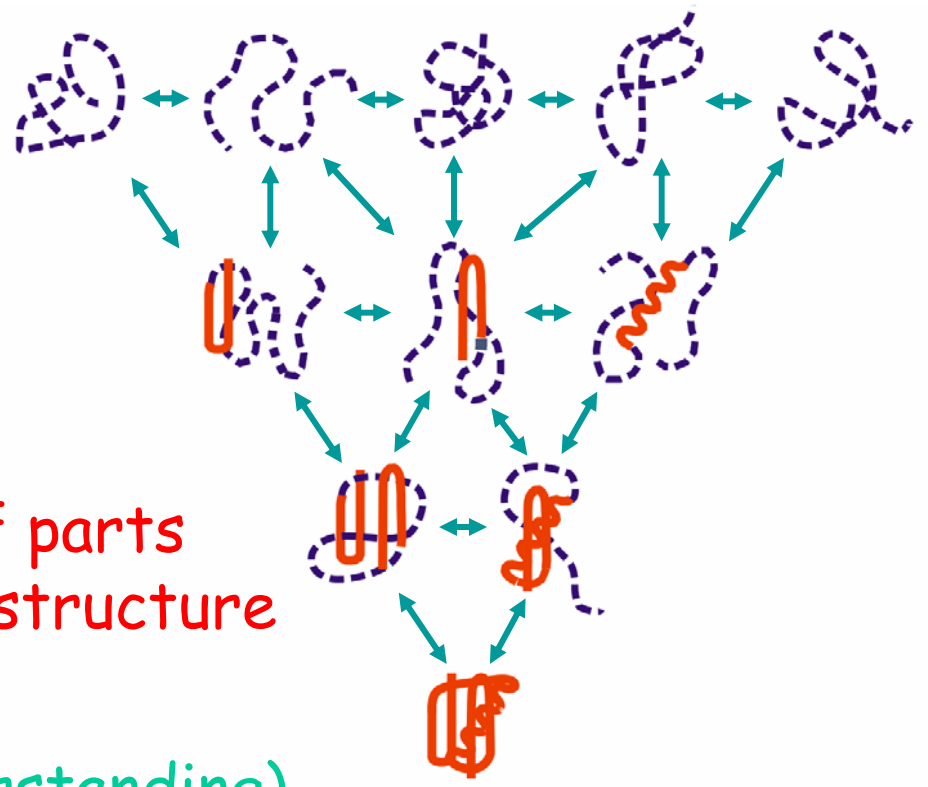
KITP
June 7, 2011

"Watching individual protein molecules fold and unfold using fluorescence spectroscopy"

(Progress toward observing transition paths)

William A. Eaton.
Laboratory of Chemical Physics
NIDDK, National Institutes of Health (NIH)
Bethesda, Maryland

The Question: ***How*** does a protein fold?

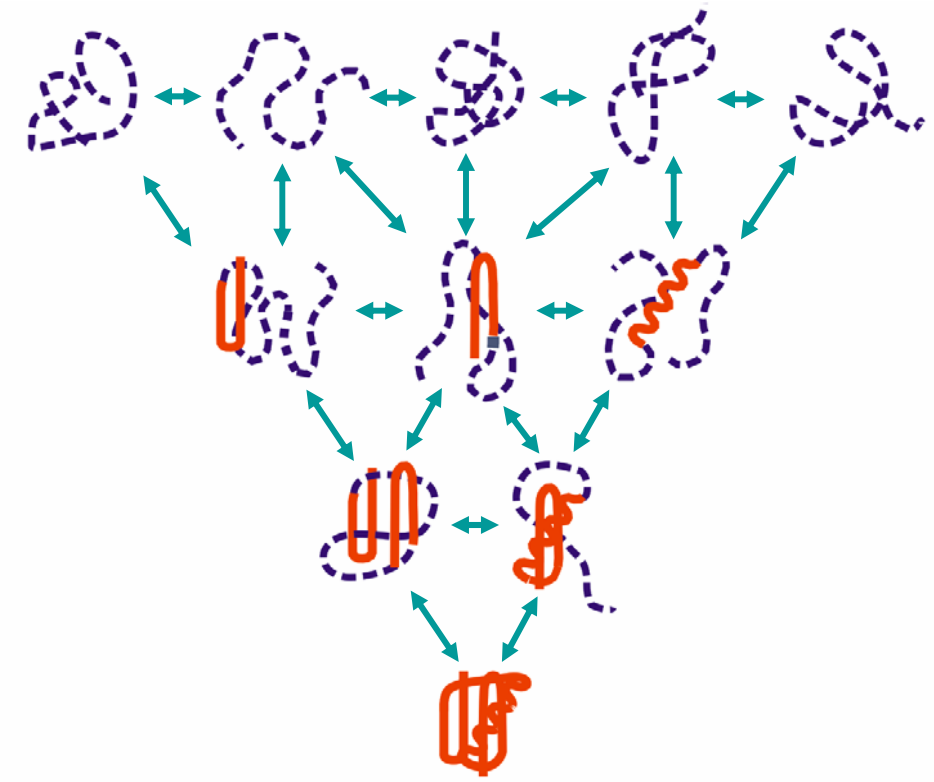


Biologist:
spontaneously!
(Anfinsen Nobel Prize)

Biochemist:
picture of order of assembly of parts
correlation between order and structure

Biophysicist (quantitative understanding)
theoretical model (i.e. partition function, master equation) - quantitative predictions of experimental measurements.
universal principles

Folding is heterogeneous with many microscopic pathways connecting the folded and unfolded states



The distribution of microscopic pathways is predicted by theory and simulations, what about experiments?

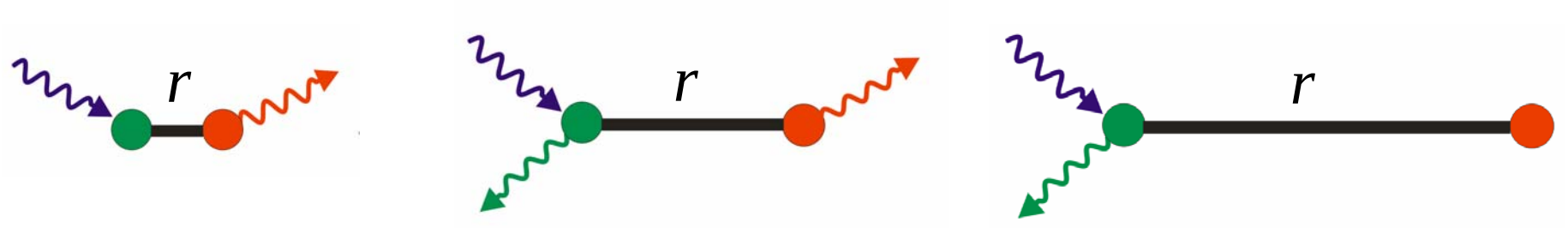
SINGLE MOLECULES!

Motivations for current experiments

Determining folding heterogeneity requires observing transition paths (a uniquely single molecule property), which is challenging and not yet observed for any system.

For an experimentalist that **can** be **fun!!**
(if it works)

Förster resonance energy transfer (FRET)

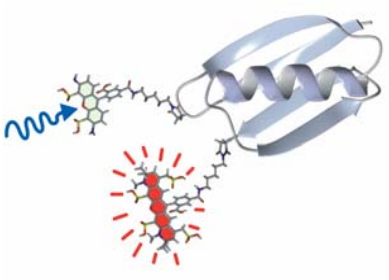


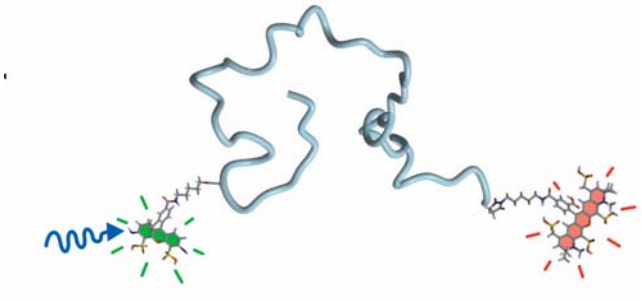
$$k_T = \frac{1}{\tau_D}\left(\frac{R_0}{r}\right)^6$$

$$\text{FRET efficiency} = \frac{k_T}{k_T + 1/\tau_D} = \frac{1}{1 + \left(r / R_0\right)^6}$$

$$= \frac{n_A}{n_A + n_D}$$

Photon trajectory for "two-state" protein



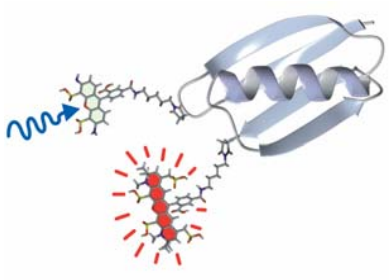


folded

time ⟶

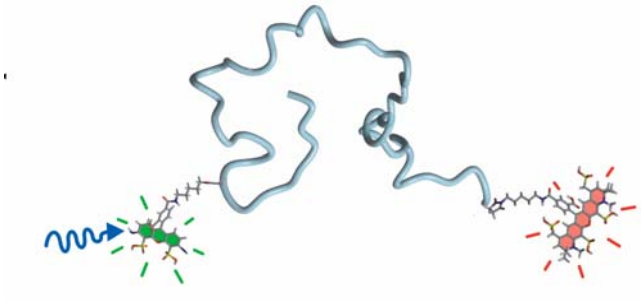
unfolded

FRET efficiency trajectory for "two-state" protein



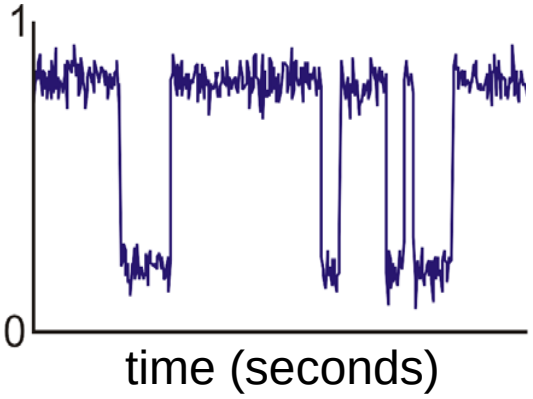
folded

time ⟶



unfolded

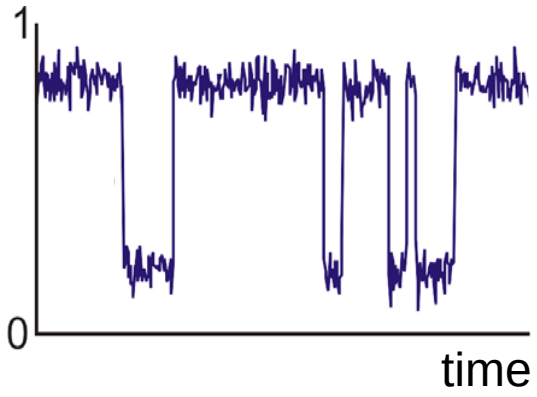
$$\text{FRET efficiency} \equiv \frac{n_A}{n_A + n_D}$$



folded

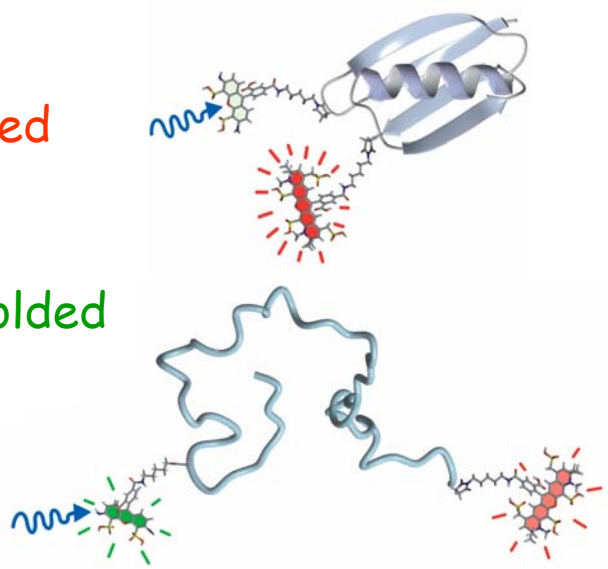
unfolded

FRET efficiency $\equiv \dfrac{n_A}{n_A + n_D}$



folded

unfolded



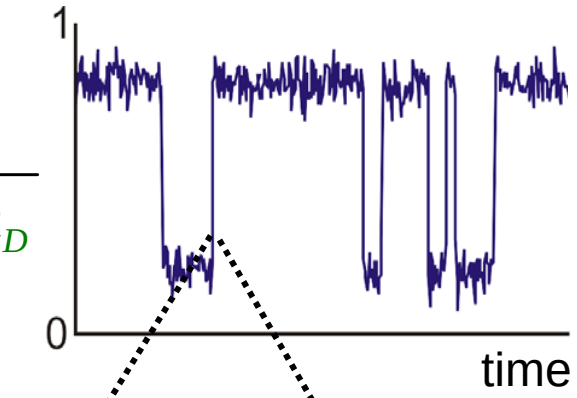
All mechanistic information contained in the "jumps":
The transition path: a property unique to single molecules

Goal for single molecule experiments: observe transition paths

Transition paths can be obtained from MD simulations for ultrafast (< 100 μs) folders, or from clever theoretical methods for slower folding proteins (e.g. H. Orland), but have never been observed ***experimentally*** for any system

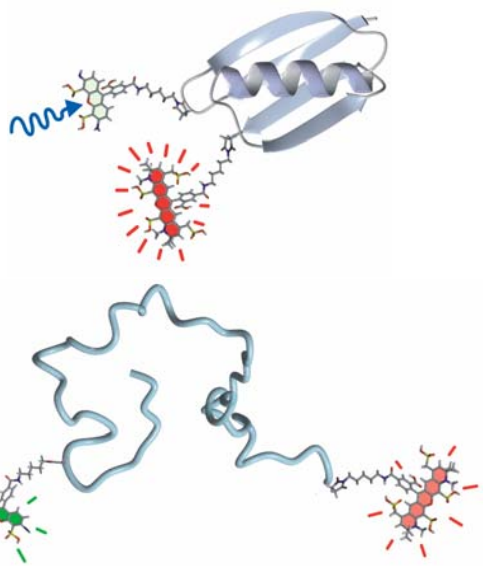
MAJOR EXPERIMENTAL CHALLENGE

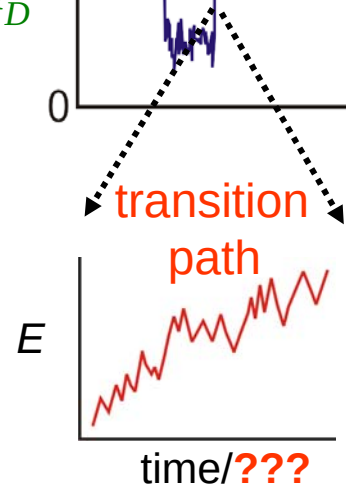
FRET efficiency $\equiv \dfrac{n_A}{n_A + n_D}$



folded

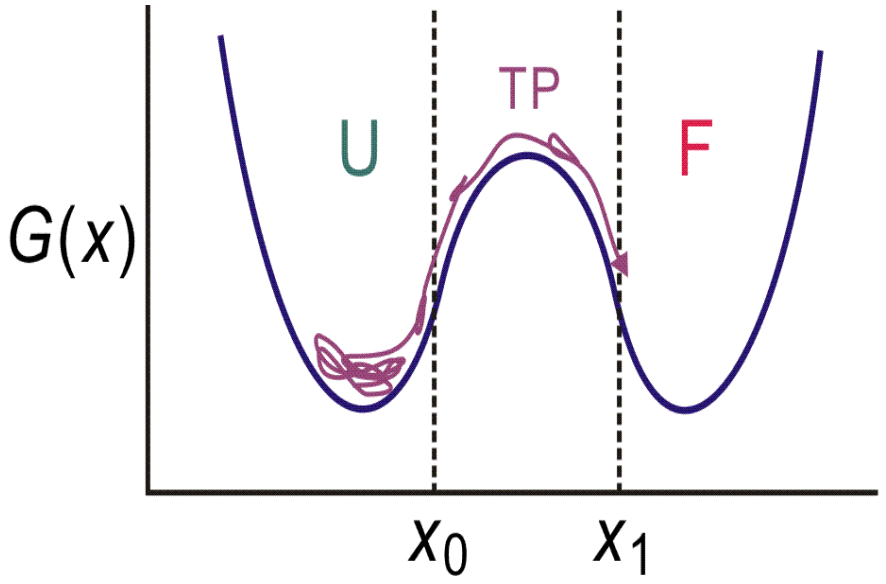
unfolded





Zeroth order question about transition paths:
what is duration?

Wolynes Energy Landscape Theory
(Socci, Onuchic, and Wolynes, JCP 1996)

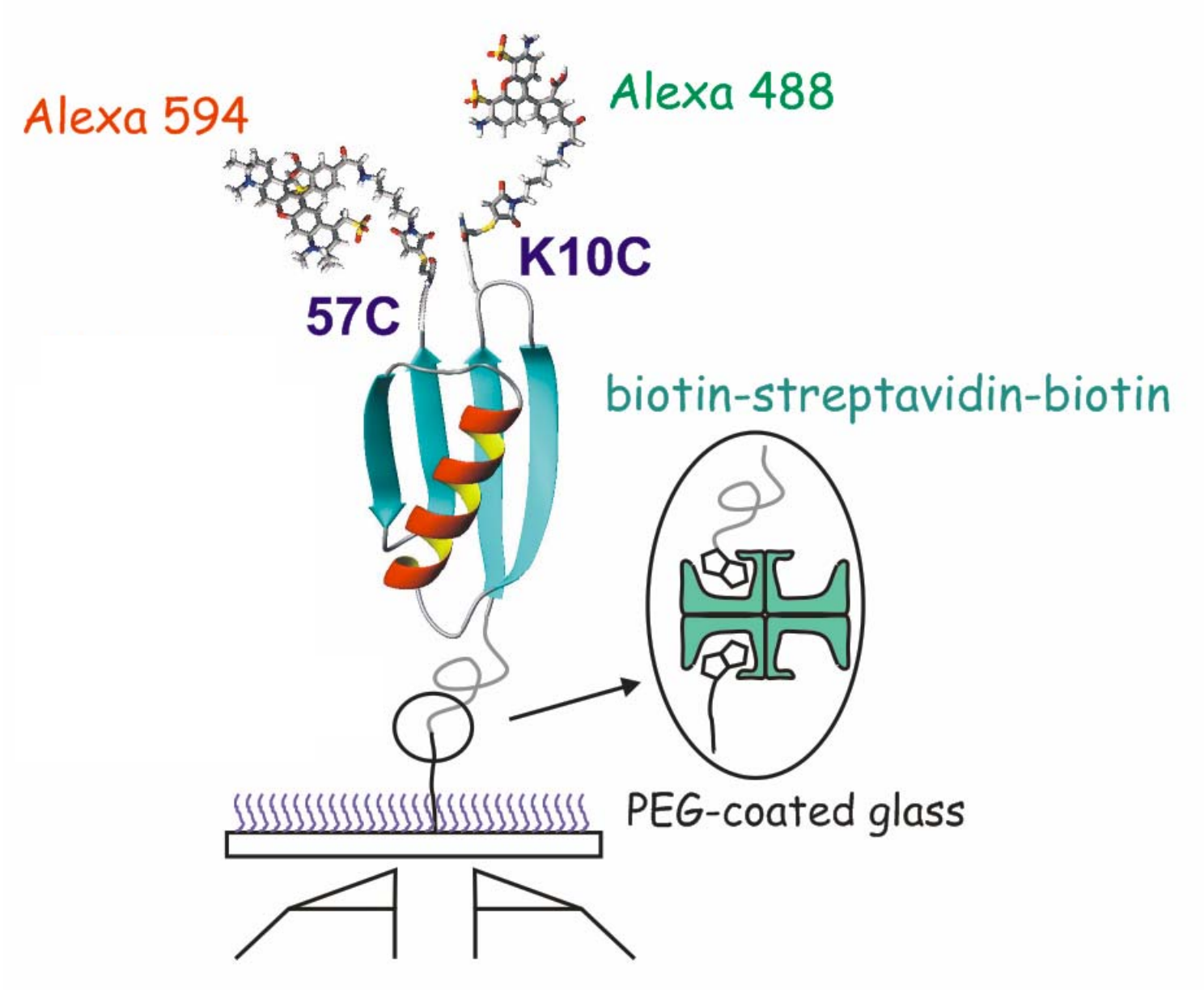


Definition of transition path time (TPT):

Trajectories that cross x_0 and reach x_1 without ever recrossing x_0

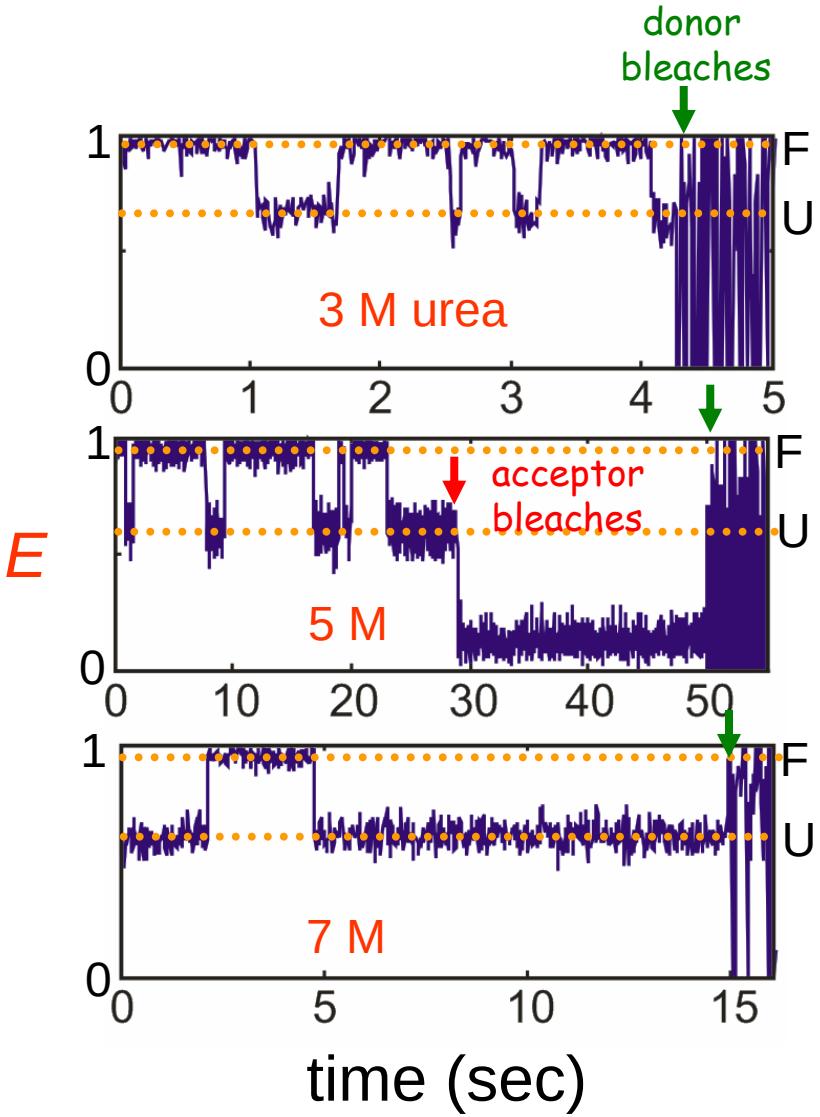
The well-studied two-state protein

56-residue protein G labeled with donor and acceptor dyes



Explained almost every one of 2,000 trajectories in detail

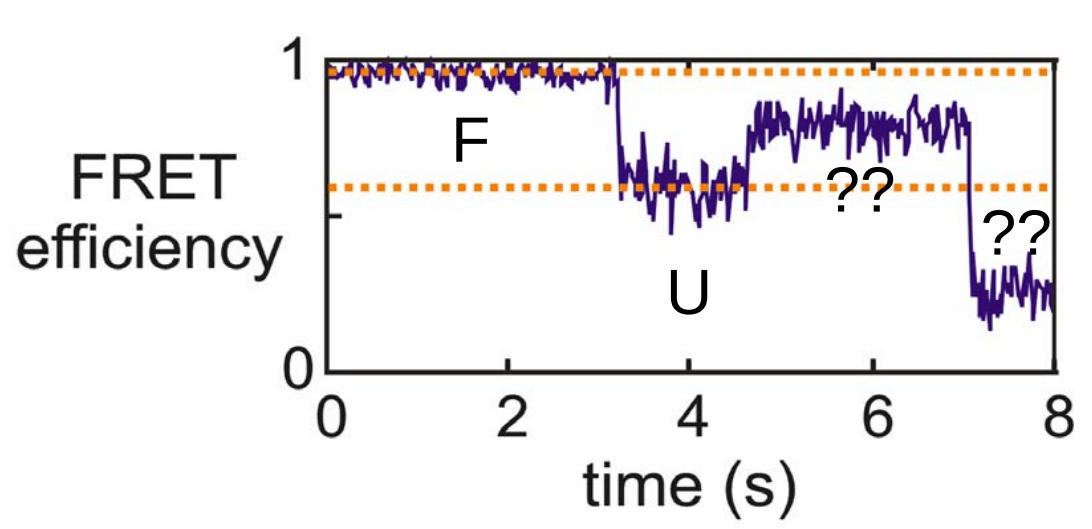
Representative (65%) folding/unfolding trajectories



Almost every trajectory explained in detail.

65% folding/unfolding
21% "blinking"
5% dye sticking to surface or linker
8% shifted donor dye spectrum

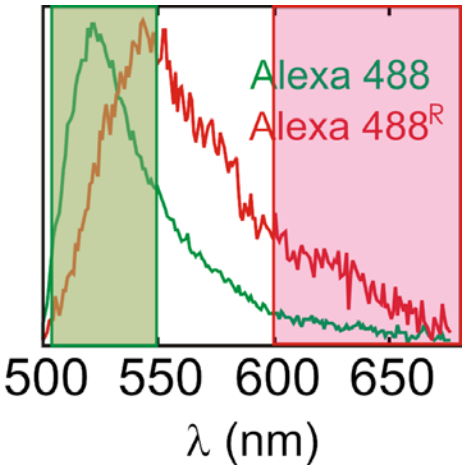
initially puzzling trajectory

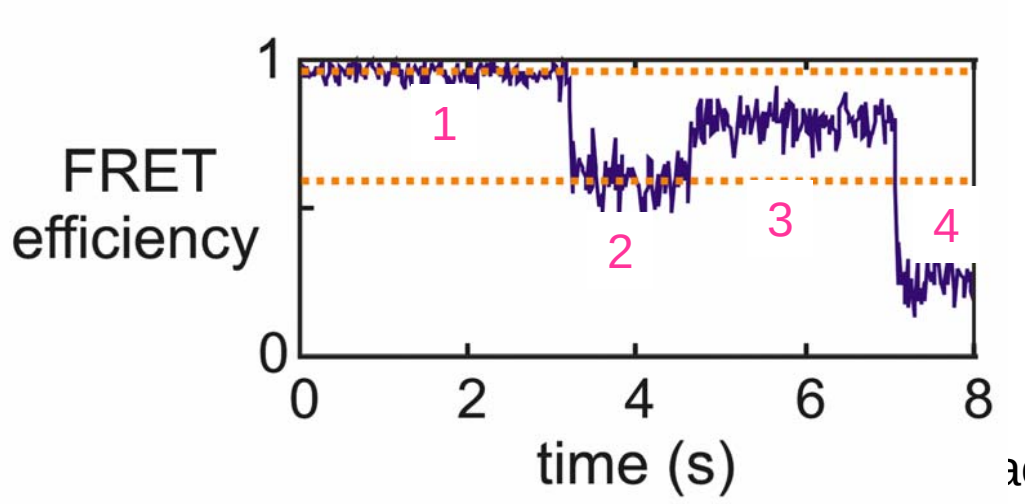


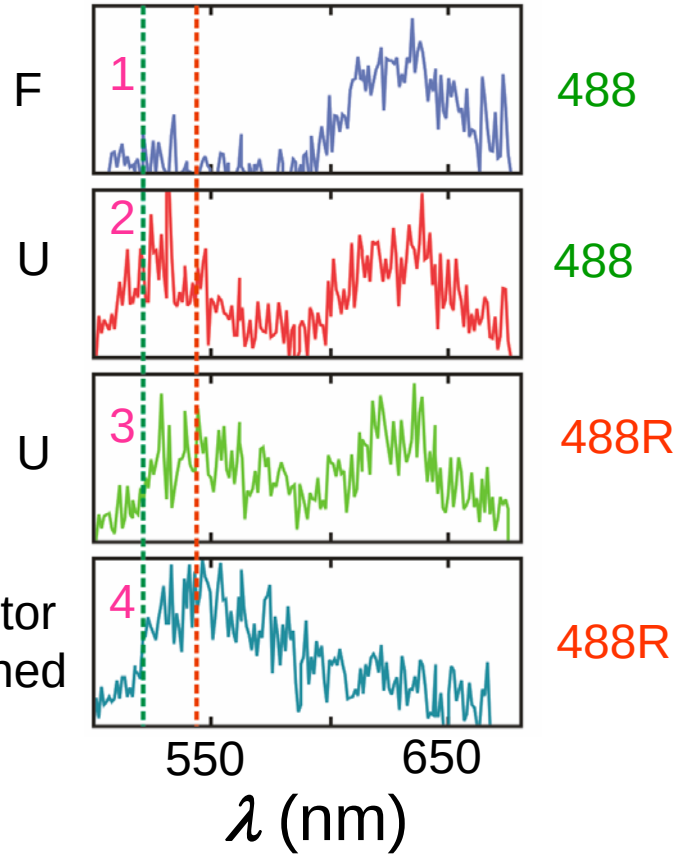
folding intermediates????

Almost every trajectory explained in detail.

65% folding/unfolding
21% "blinking"
5% dye sticking to surface or linker
8% shifted donor dye spectrum

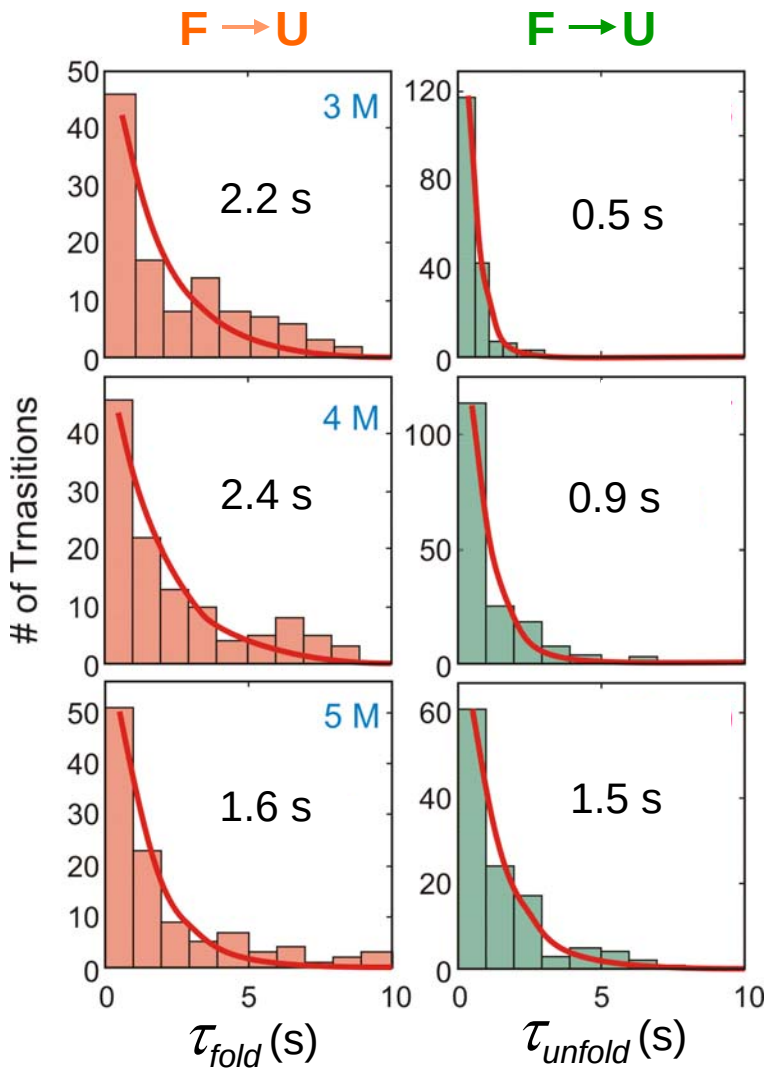




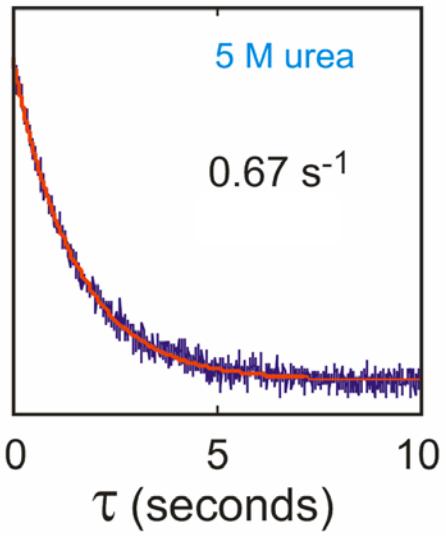


Comparison of Kinetics

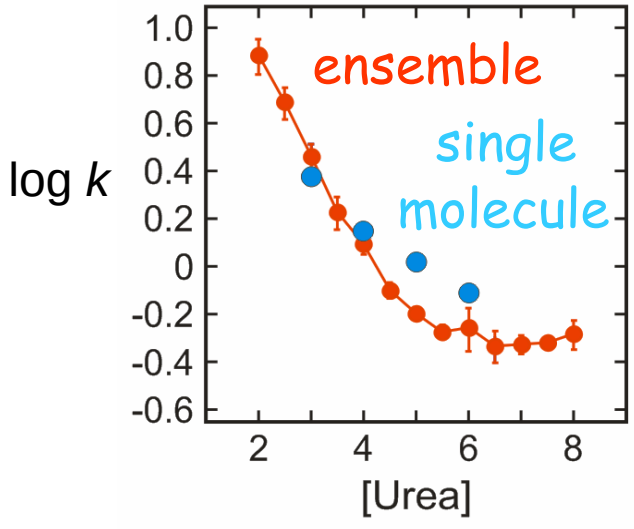
Waiting times are exponentially distributed



stopped-flow

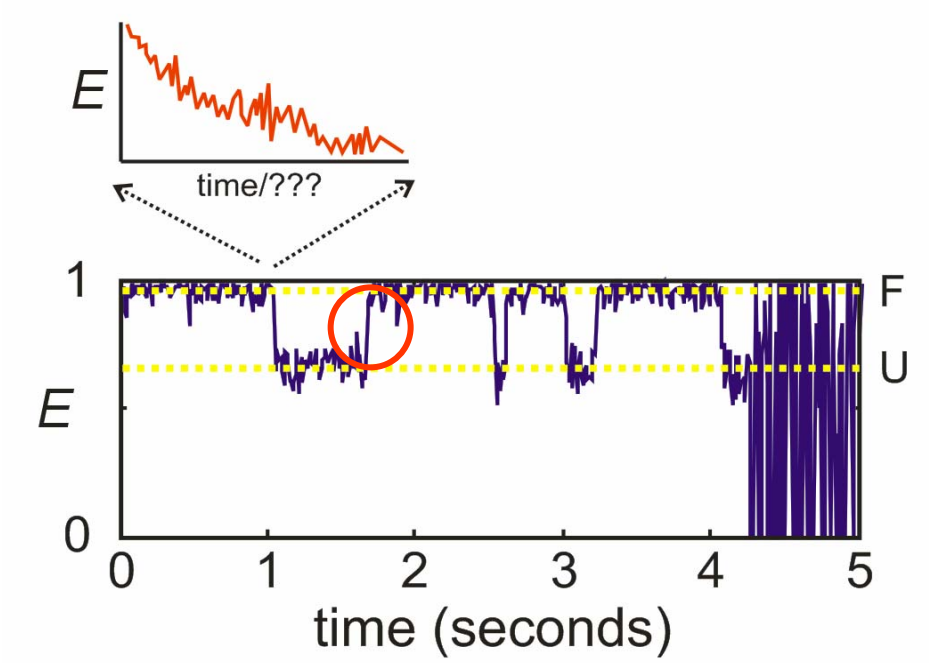


relaxation rate, $k = k(U \rightarrow F) + k(F \rightarrow U)$

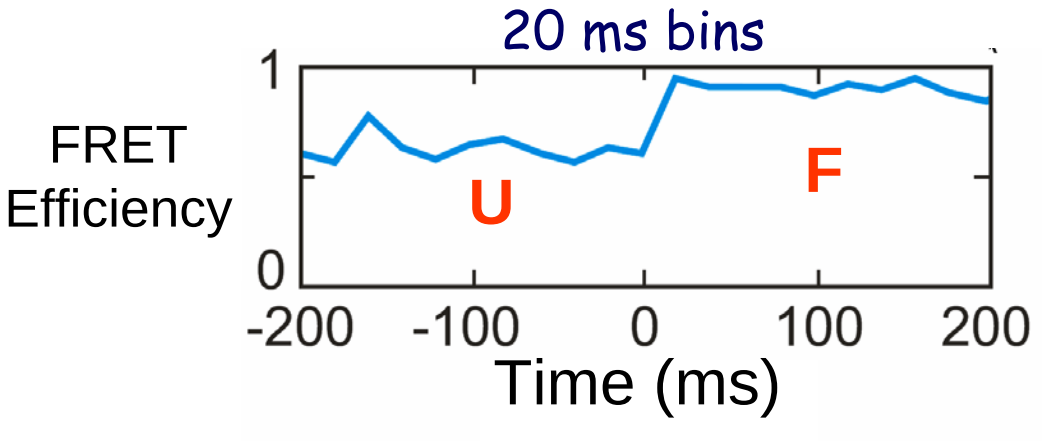


Small effect of immobilization

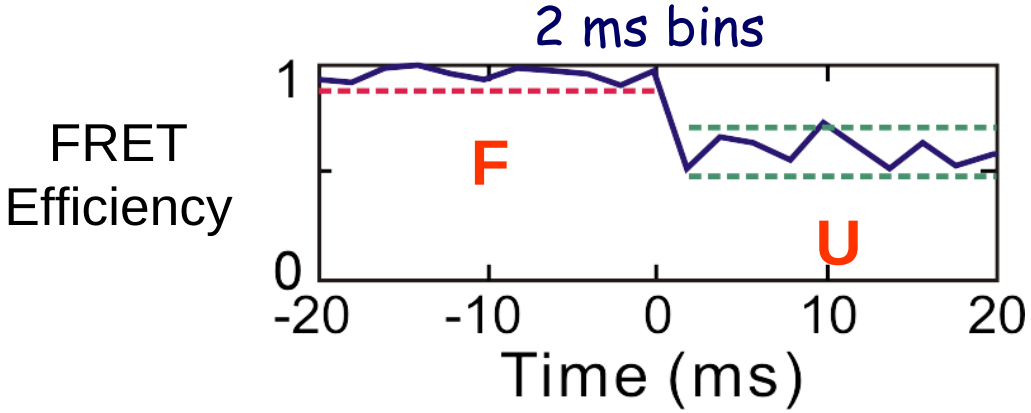
All mechanistic information contained in "jumps" (transition paths)



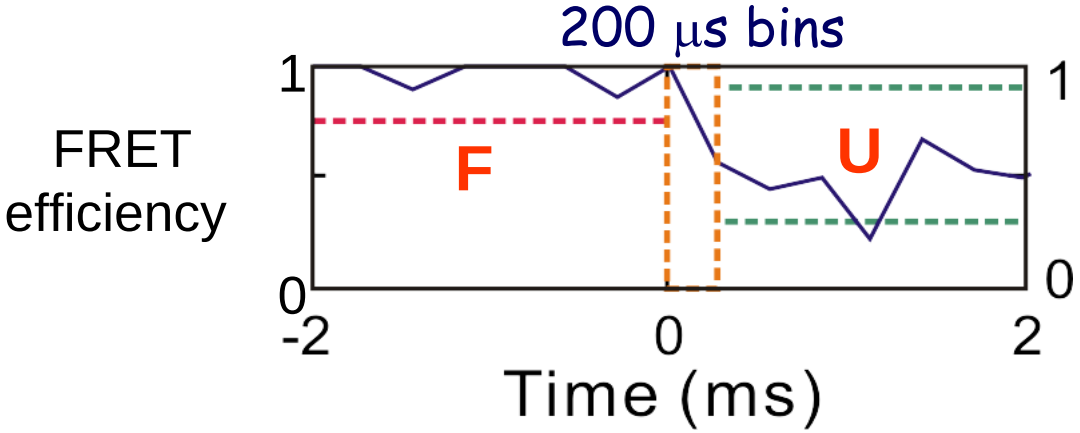
What can we say about transition path times?



5 photons/ms
Bleach in ~10 s
~1000 transitions
~2000 trajectories



50 photons/ms
Bleach in ~0.1 s
46 transitions in
1500 trajectories



Transition still in single bin, ∴
TPT < 200 μs

TPT < 200 μs (>10,000x shorter than folding time)
(Chung, Louis, & Eaton, *PNAS* 2009)

Can we measure a transition path time?

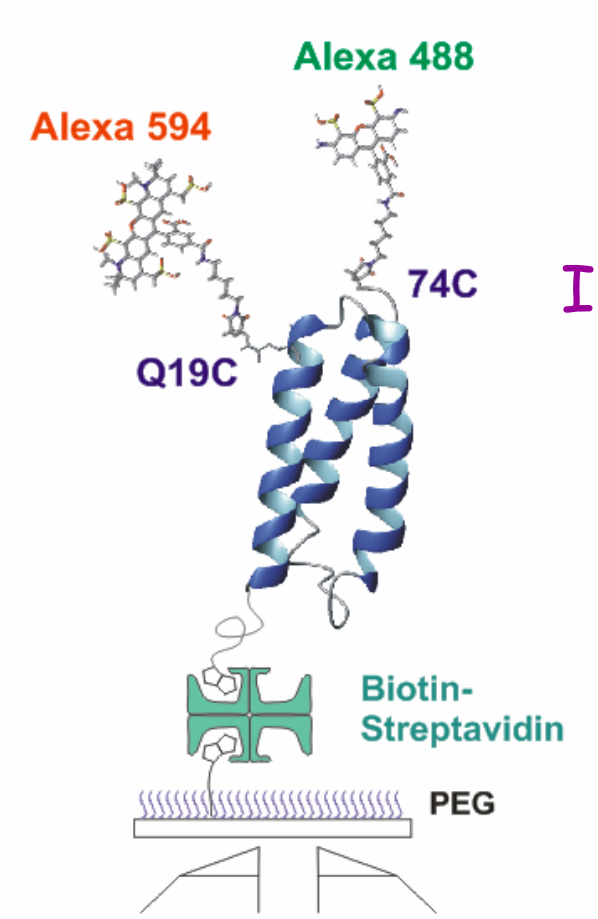
YES, but we will have to

more intense excitation to increase photons
(shortens bleaching time)

observe a very large number of transitions
(automate data acquisition, faster folding protein)

analyze these transitions ***collectively***
(maximum likelihood method)

Chung et al., J Phys Chem A 2011

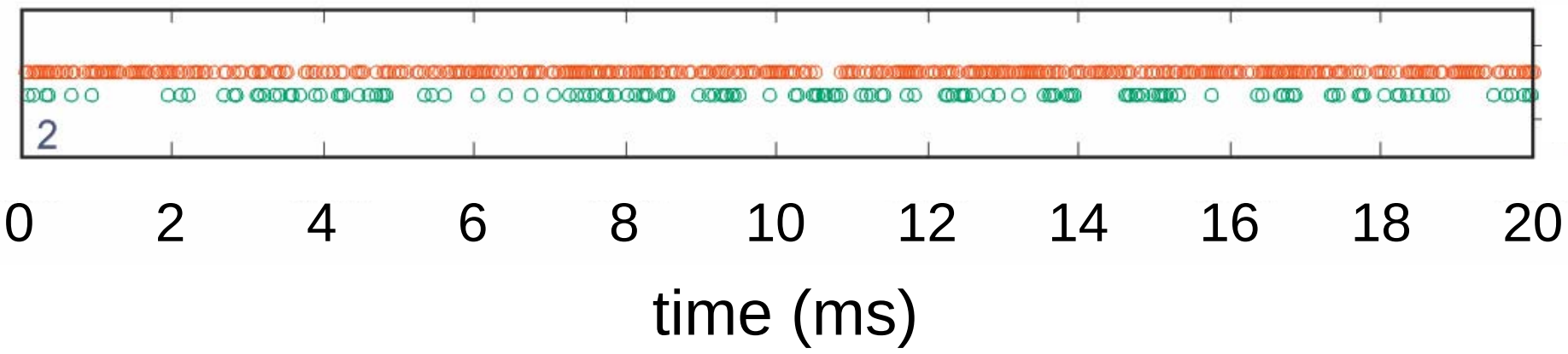


α_3D, two-state folder, τ_F = 1 ms

Instead of 1 transition/2s for protein G,
1000 transitions/s for α_3D

~1,000 transitions at each denaturant concentration, ~ 50 photons/ms,
~30 ms bleaching time

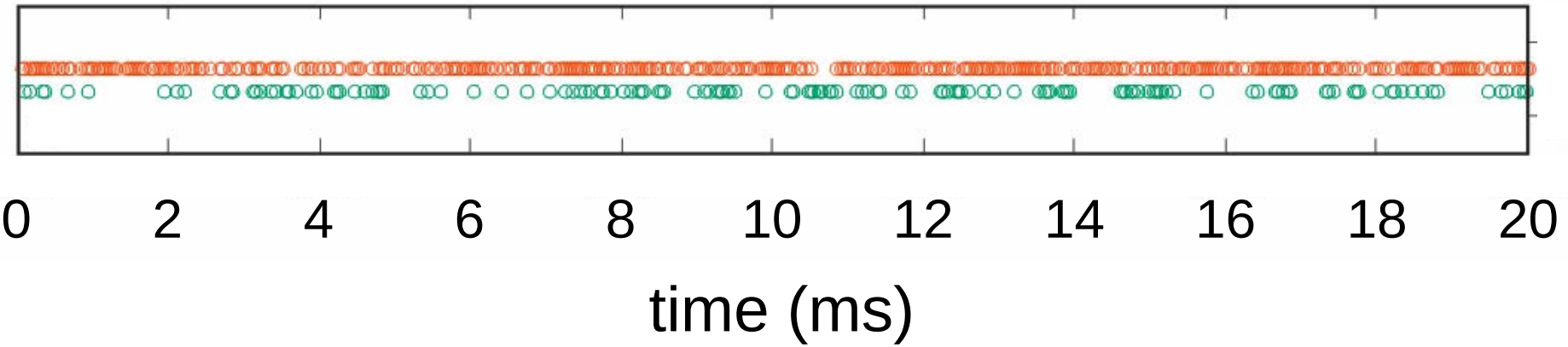
photon trajectory



Gopich/Szabo Maximum Likelihood Function

(J Phys Chem B, 2009)

Given a model, the method yields the most likely model parameters consistent with the photon trajectories

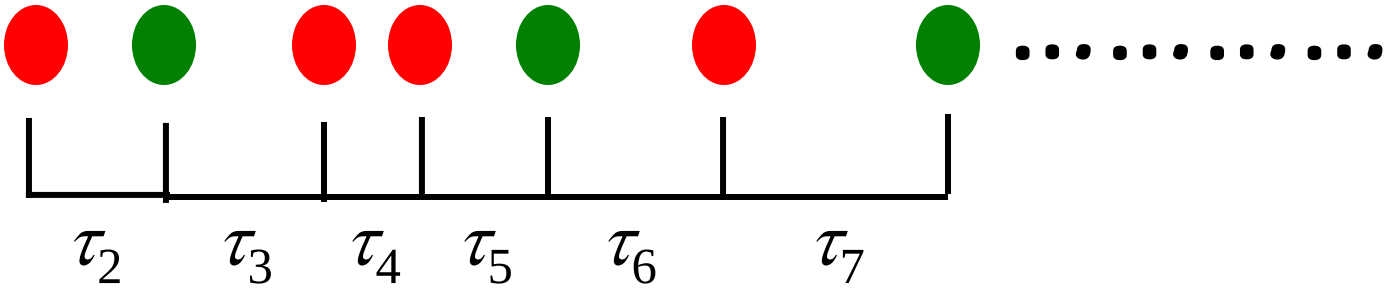


Gopich/Szabo Maximum Likelihood Function for Two-State Model

$$\mathrm{U} \underset{k_{\mathrm{U}}}{\overset{k_{\mathrm{F}}}{\rightleftharpoons}} \mathrm{F}$$

$$L_j = (1 \;\; 0) \prod_{k=2}^{N_j} \left[\mathbf{\Phi}(c_k) \begin{pmatrix} 1 & 0 \\ 0 & e^{-(k_{\mathrm{F}}+k_{\mathrm{U}})\tau_k} \end{pmatrix} \right] \mathbf{\Phi}(c_1) \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

- $\Phi(\text{acceptor}) = \begin{pmatrix} \varepsilon_{\mathrm{F}} p_{\mathrm{F}} + \varepsilon_{\mathrm{U}} p_{\mathrm{U}} & (\varepsilon_{\mathrm{U}} - \varepsilon_{\mathrm{F}}) p_{\mathrm{U}} \\ (\varepsilon_{\mathrm{U}} - \varepsilon_{\mathrm{F}}) p_{\mathrm{F}} & \varepsilon_{\mathrm{F}} p_{\mathrm{U}} + \varepsilon_{\mathrm{U}} p_{\mathrm{F}} \end{pmatrix}$
- $\Phi(\text{donor}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \Phi(\text{acceptor})$



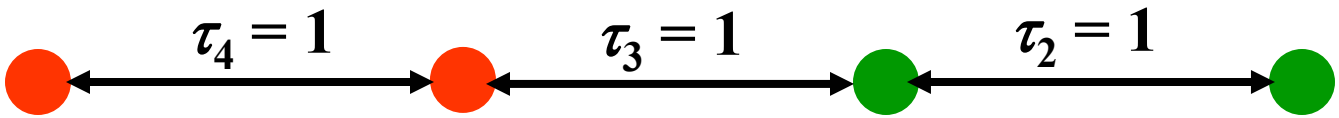
$$\mathrm{U} \underset{k_\mathrm{U}}{\overset{k_\mathrm{F}}{\rightleftharpoons}} \mathrm{F}$$

$$L_j = (1 \;\; 0)\prod_{k=2}^{N_j}\left[\boldsymbol{\Phi}(c_k)\begin{pmatrix}1 & 0\\ 0 & e^{-(k_\mathrm{F}+k_\mathrm{U})\tau_k}\end{pmatrix}\right]\boldsymbol{\Phi}(c_1)\begin{pmatrix}1\\0\end{pmatrix}$$

$$\Phi(\text{acceptor}) = \begin{pmatrix}\varepsilon_\mathrm{F}p_\mathrm{F} + \varepsilon_\mathrm{U}p_\mathrm{U} & (\varepsilon_\mathrm{U}-\varepsilon_\mathrm{F})p_\mathrm{U}\\ (\varepsilon_\mathrm{U}-\varepsilon_\mathrm{F})p_\mathrm{F} & \varepsilon_\mathrm{F}p_\mathrm{U} + \varepsilon_\mathrm{U}p_\mathrm{F}\end{pmatrix}$$

$$\Phi(\text{donor}) = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} - \Phi(\text{acceptor})$$

$p_\mathrm{U} = p_\mathrm{F} = 1/2,\;\; \varepsilon_\mathrm{U} = 0,\; \varepsilon_\mathrm{F} = 1,\; \boldsymbol{\tau_k = 1},\;\; \boldsymbol{k_\mathrm{U} = k_\mathrm{F} = ??}$

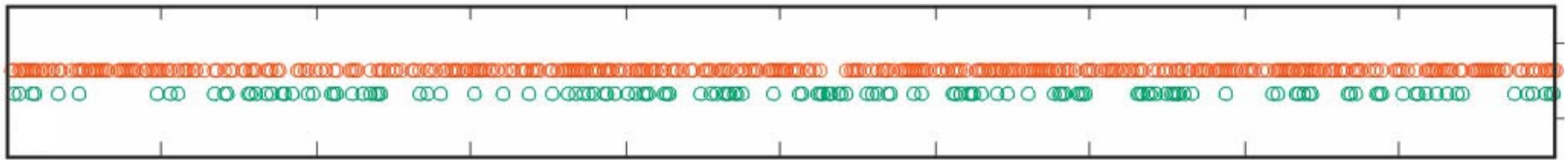


$$L(k_\mathrm{F} = 0.5) = (1\;0)\left[\begin{pmatrix}1/2 & -1/2\\ -1/2 & 1/2\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1/e\end{pmatrix}\right]\left[\begin{pmatrix}1/2 & -1/2\\ -1/2 & 1/2\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1/e\end{pmatrix}\right]\left[\begin{pmatrix}1/2 & 1/2\\ 1/2 & 1/2\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1/e\end{pmatrix}\right]\begin{pmatrix}1/2 & 1/2\\ 1/2 & 1/2\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}$$

$= 0.074$

k_F	L
0.25	0.063
1	0.070

photon-by-photon analysis to obtain k_F, k_U, ε_F, ε_U



0 2 4 6 8 10 12 14 16 18 20

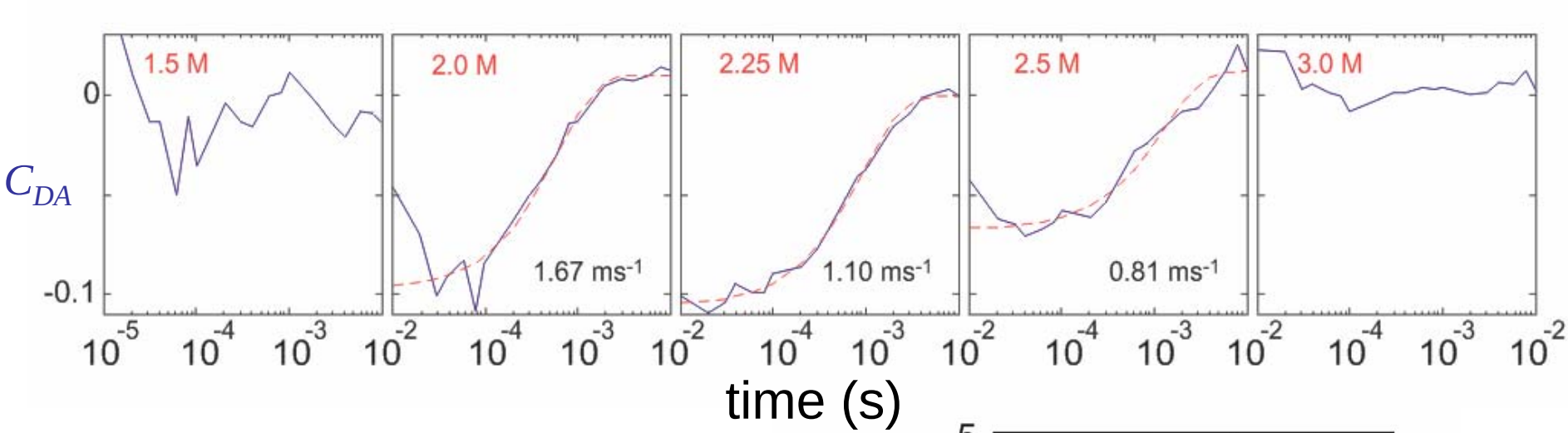
time (ms)

GdmCl (M)	ε_F	ε_U	k (ms^{-1})	p_F
1.5	0.94	0.64	2.64 (±0.23)	0.92
2	0.93	0.64	1.29 (±0.05)	0.67 (±0.01)
2.25	0.93	0.61	0.99 (±0.04)	0.43 (±0.01)
2.5	0.93	0.62	1.06 (±0.04)	0.32
3	0.92	0.56	1.22 (±0.08)	0.05

Are the results accurate?

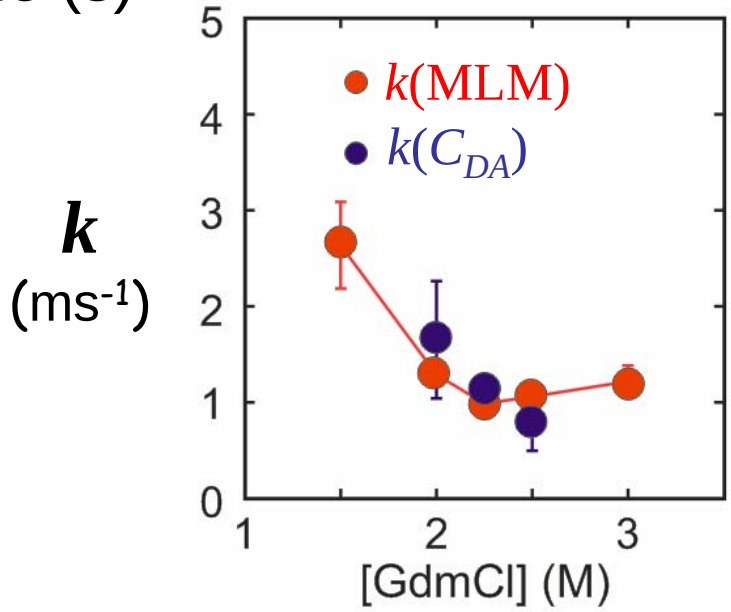
Compare with the donor-acceptor cross-correlation function decay

$$C_{DA}(\tau) = \frac{\langle n_D(t) n_A(t+\tau) \rangle}{\langle n_D(t) \rangle \langle n_A(t) \rangle} - 1$$



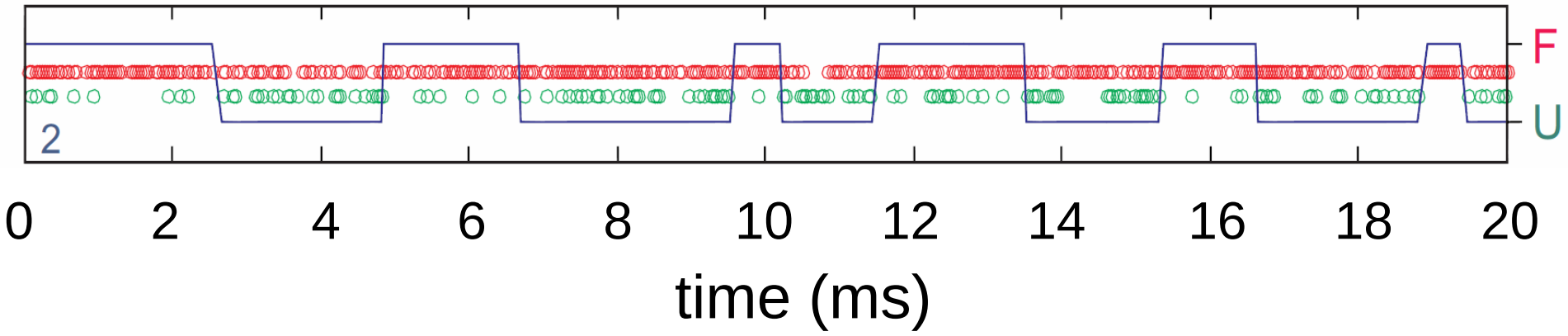
$$k(C_{DA}) = k(\mathrm{MLM})$$

$$k(\mathrm{MLM}) = k_F(\mathrm{MLM}) + k_U(\mathrm{MLM})$$



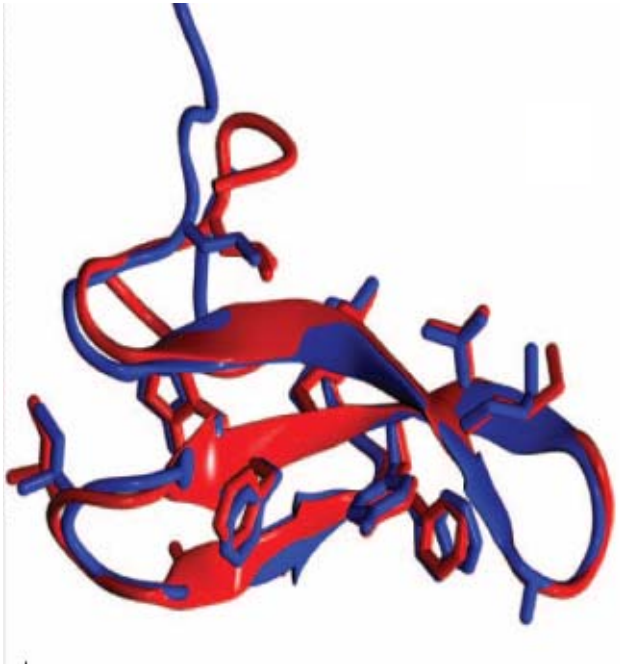
Strategy

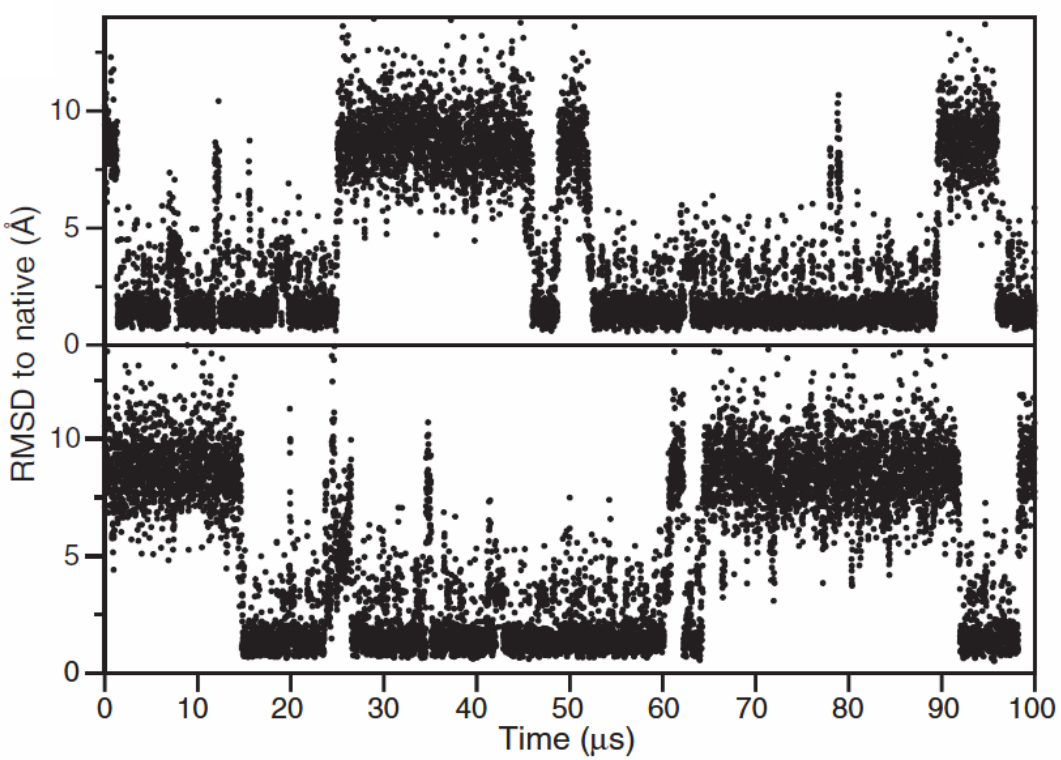
Apply Viterbi algorithm to produce most probable state trajectory to locate transition region



and only analyze the trajectory in the vicinity of the transition

David E. Shaw et al.: *Science* October 15, 2010 issue, fully atomistic MD simulation of ultrafast folder using "Anton", hard-wired computer for MD calculations
- 35 residue WW domain





simulation: τ_F = 10 ± 3 μs, experiment: τ_F = 14 ± 2 μs
<TPT> ≈ 2 μs at 20°C (after viscosity correction)

We now have a good estimate of the average transition path time for an ultrafast (10 microsecond)-folding protein

What can we say about our 2 second-folding protein G – inaccessible even to "Anton."

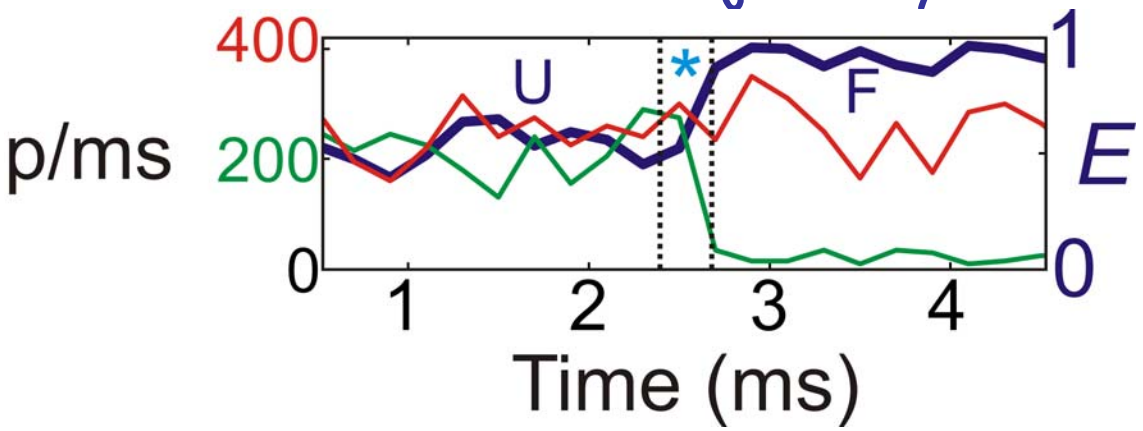
Hoi Sung got ambitious!

Measured 46,932 trajectories for protein G, τ_F = 2 s

350 photons/ms; bleaching time ~ 10 ms

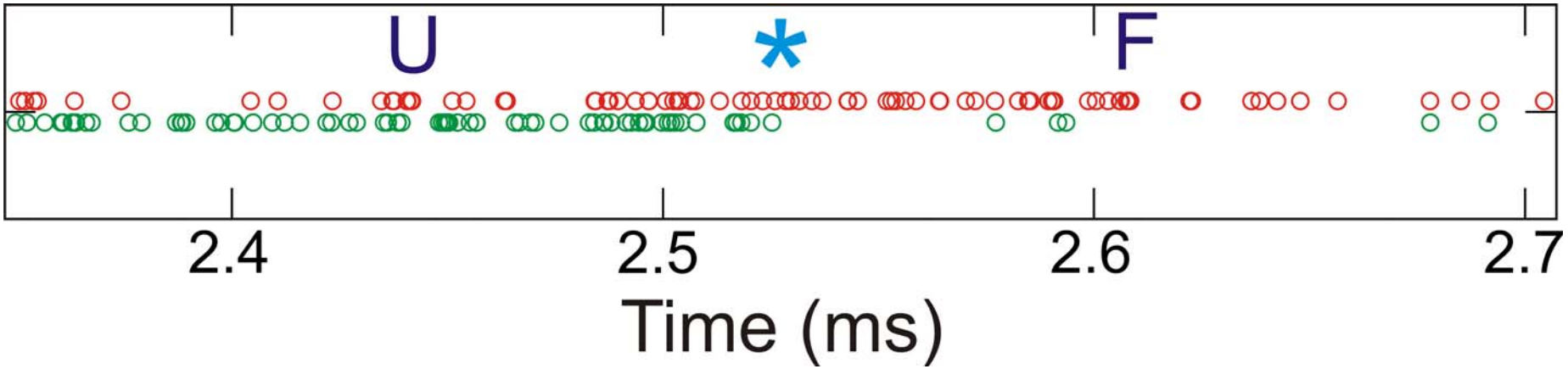
observed 151 transitions

FRET trajectory



200 μs bins

Photon trajectory

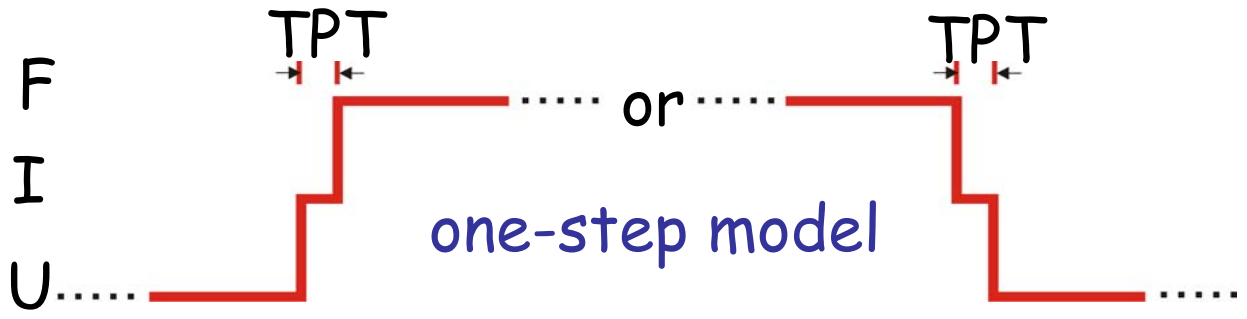


Simulate trajectories with model to answer question:

Given the current data (number of transitions, count rate, FRET efficiencies,), how short a transition path time we can expect to be able to determine with the current data ?

151 transitions, 350 photons/ms, $\varepsilon_F = 0.95$ $\varepsilon_U = 0.60$

Simulations of trajectories for one-step model:



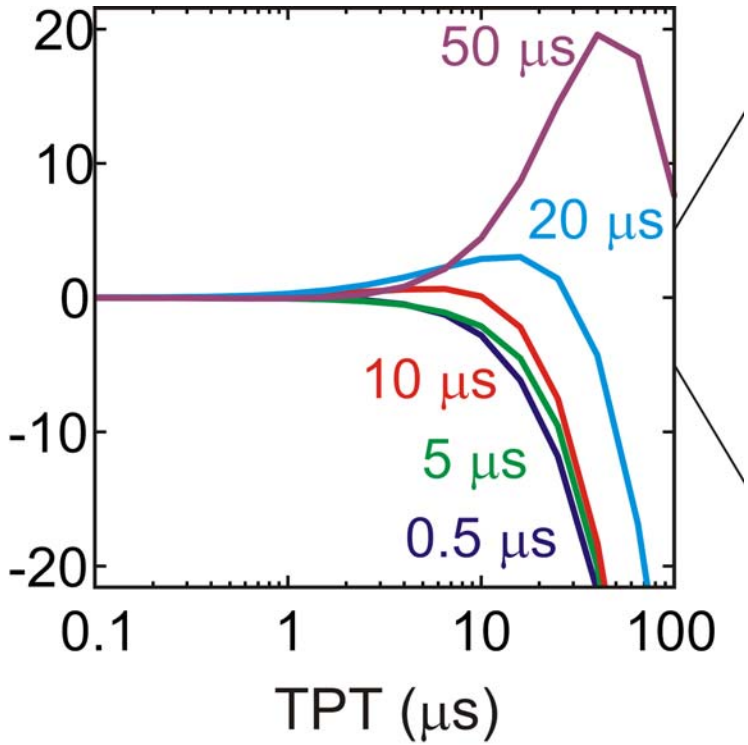
$$\langle TPT \rangle = 1/k_{\mathrm{I}\to\mathrm{F}} = 1/k_{\mathrm{I}\to\mathrm{U}} \qquad \varepsilon_{\mathrm{I}} = \frac{\varepsilon_{\mathrm{U}} + \varepsilon_{\mathrm{F}}}{2}$$

$$\mathbf{K} = \begin{pmatrix} -k_{\mathrm{U}\to\mathrm{I}} & k_{\mathrm{I}\to\mathrm{U}} & 0 \\ k_{\mathrm{U}\to\mathrm{I}} & -(k_{\mathrm{I}\to\mathrm{F}} + k_{\mathrm{I}\to\mathrm{U}}) & k_{\mathrm{F}\to\mathrm{I}} \\ 0 & k_{\mathrm{I}\to\mathrm{F}} & -k_{\mathrm{F}\to\mathrm{I}} \end{pmatrix}$$

Analyse simulated trajectories using Gopich/Szabo function with and without transition path, and compare likelihoods

(TPT is only variable parameter in likelihood function)

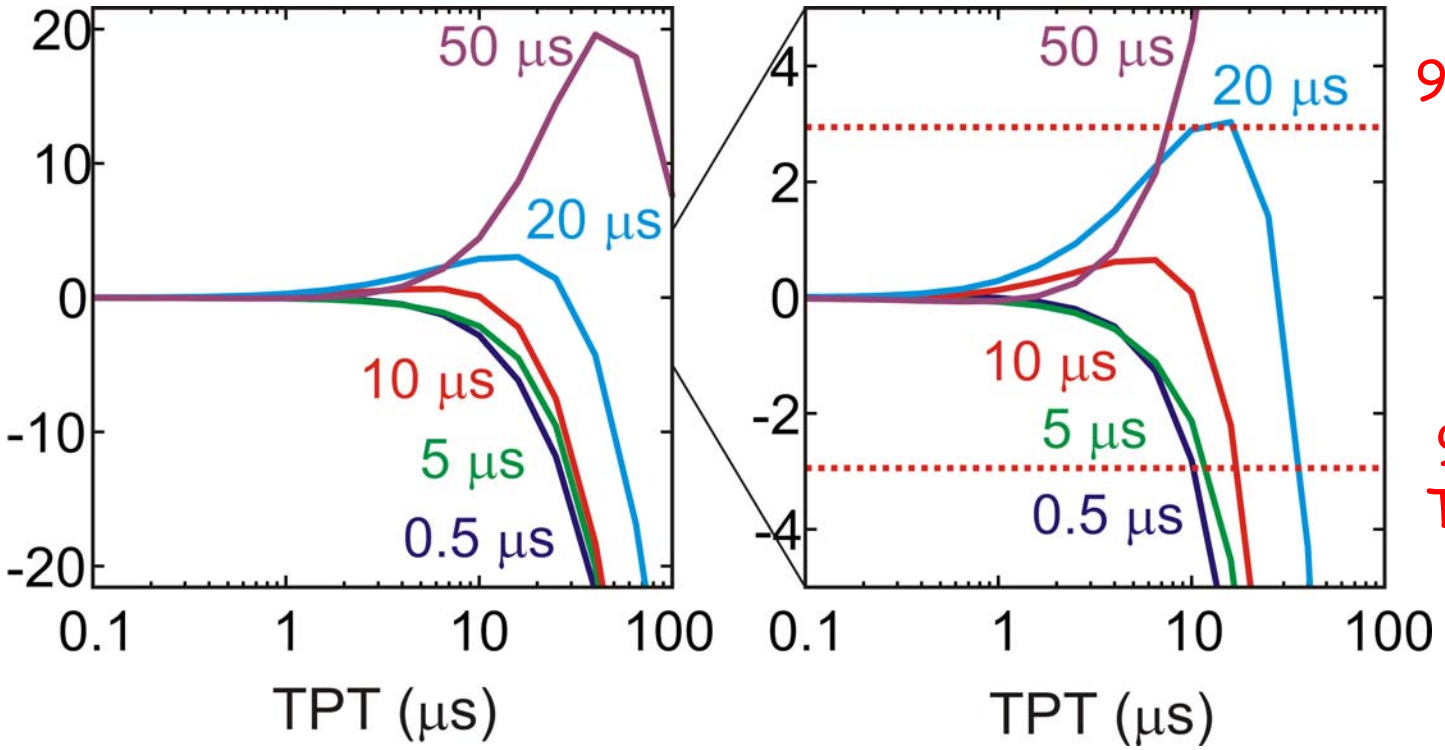
$\ln L(\mathrm{TPT}) - \ln L(0)$



Analyse simulated trajectories using Gopich/Szabo function with and without transition path, and compare likelihoods

(TPT is only variable parameter in likelihood function)

ln L(TPT) – ln L(0)

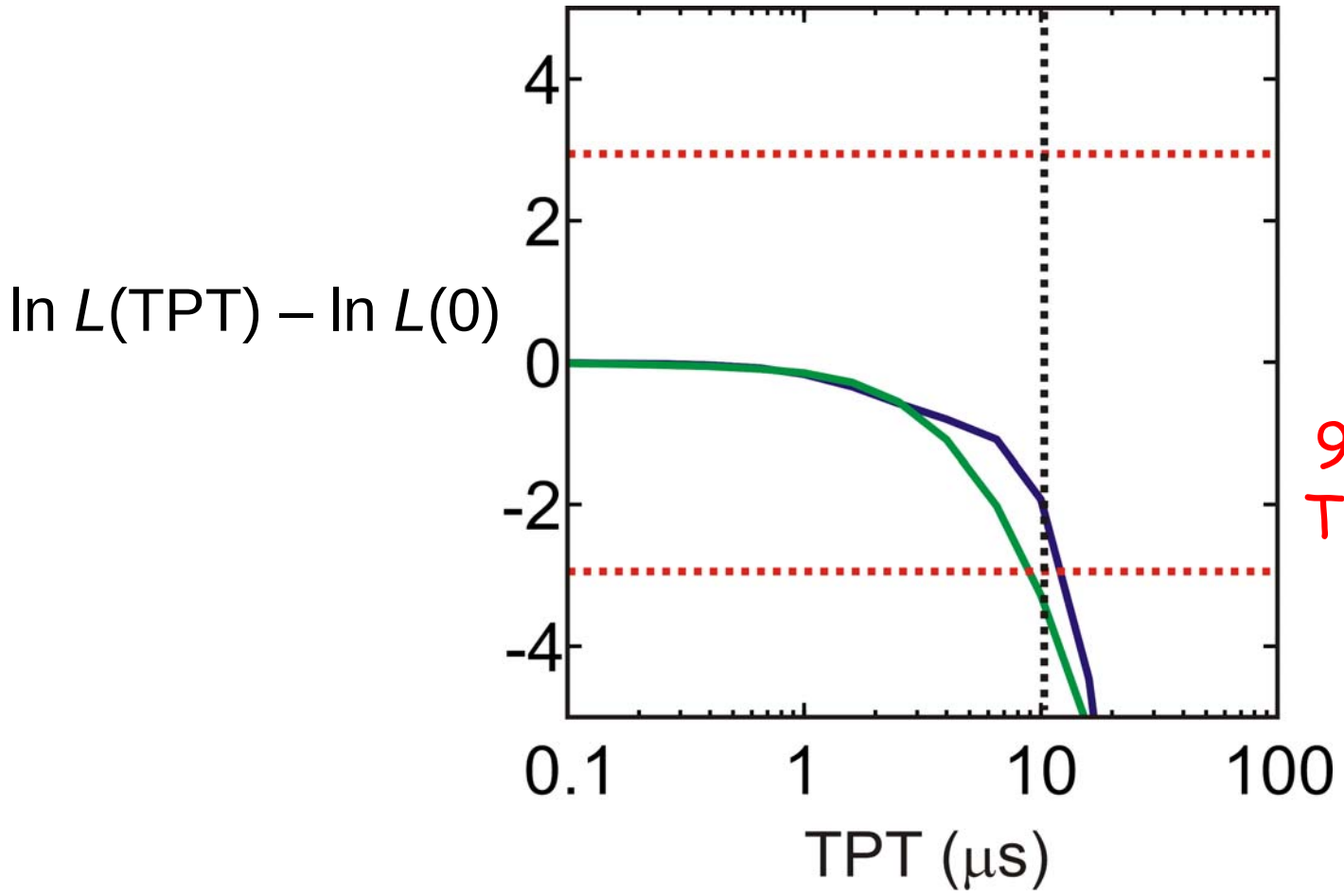


95% confidence peak at TPT

95% confidence TPT upper bound

Experiment

111 unfolding transitions and 40 folding transitions for Protein G with folding time of ~ 2 s



95% confidence TPT upper bound **~ 10 μs**

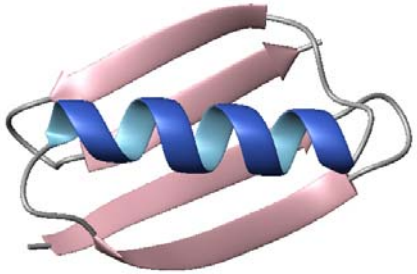
TPT 200,000-fold < 2 s folding time

WW domain (Shaw, Gruebele)

$\langle \tau_F \rangle$ = 10 μs

<TPT> ~ 2 μs

Protein G



$\langle \tau_F \rangle$ = 2 s

<TPT> < 10 μs

Transition path times differ by less than 5-fold for folding times that differ by 200,000-fold

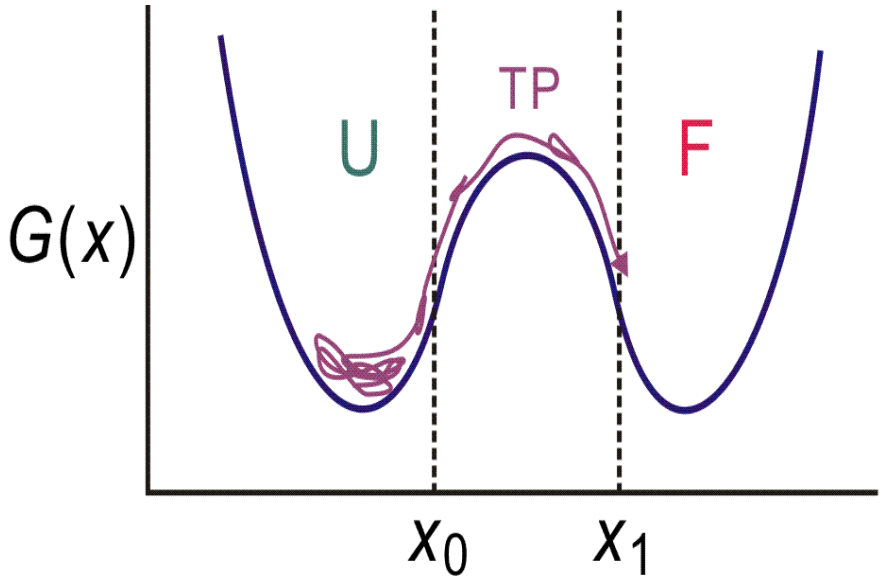
Conclusion (with caveats)

A slow folder and an ultrafast folder take almost the same time to fold when it actually happens!!

Transition path times differ by less than 5-fold for folding times that differ by 200,000-fold

How do we explain this result?

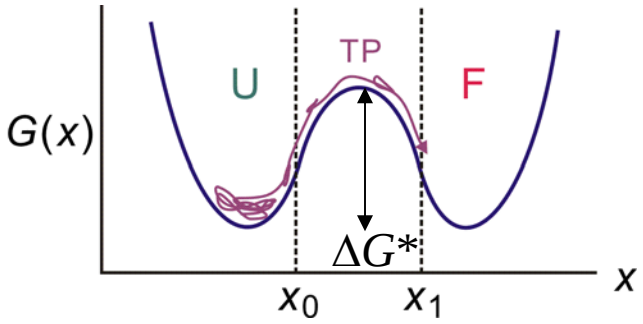
Wolynes Energy Landscape Theory
(Socci, Onuchic, and Wolynes, JCP 1996)



Definition of transition path time (TPT):

Duration of trajectory that cross x_0 and reaches x_1 without ever recrossing x_0

Prediction from energy landscape theory



$\beta = 1/k_B T$
ΔG^*: barrier height
ω^2: well curvature
$(\omega^*)^2$: barrier curvature
D: diffusion coefficient

$$1/k_F = \langle \tau_F \rangle = \tau_0 \exp\left(\beta \Delta G_F^*\right) \overset{\text{Kramers}}{=} \frac{2\pi}{D^* \beta \omega \omega^*} \exp\left(\beta \Delta G_F^*\right)$$

For diffusive crossing of harmonic barrier from x_0 to x_1,

$$\langle TPT \rangle \overset{\text{Szabo}}{\approx} \frac{\ln\left(3\beta \Delta G^*\right)}{D^* \beta (\omega^*)^2}$$

For $\omega\omega^* = (\omega^*)^2$

$$\langle TPT \rangle \approx \tau_0 \frac{\ln\left(3\beta \Delta G^*\right)}{2\pi}, \quad \langle TPT \rangle \approx \tau_0 \frac{\ln\left(3\ln(\tau_F/\tau_0)\right)}{2\pi}$$

Transition path time is insensitive to barrier height

Experiment

$$\frac{\tau_{\text{F, protein G}}}{\tau_{\text{F, WW domain}}} = 200{,}000 \quad \frac{\langle TPT \rangle_{\text{protein G}}}{\langle TPT \rangle_{\text{WW domain}}} < 5$$

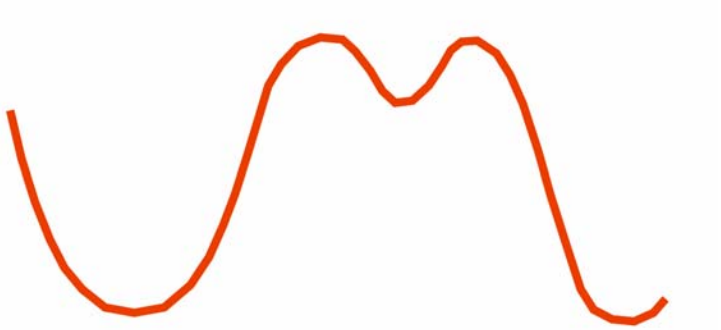
Landscape Theory (same D, ω, ω^*)

($\tau_0 \sim 1\ \mu s$)

$$\frac{\langle TPT \rangle_{\text{protein G}}}{\langle TPT \rangle_{\text{WW domain}}} \approx \frac{\ln\left(3\ln\left(\tau_{\text{protein G}} / \tau_0\right)\right)}{\ln\left(3\ln\left(\tau_{\text{WW domain}} / \tau_0\right)\right)} \approx 2$$

Caveats to theoretical estimate

1. High-lying free energy minima will slow TPT



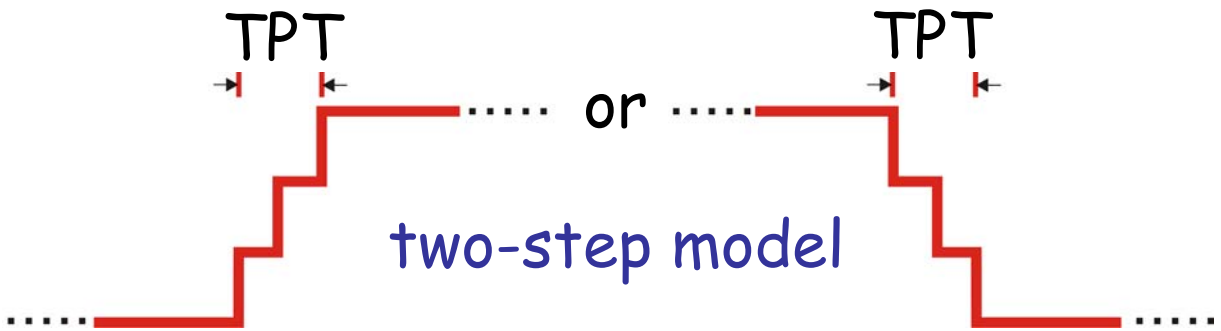
2. Very different diffusion coefficients, i.e. much rougher underlying energy landscape in slow folders

Caveat to experiment

FRET may only monitor part of the transition path.

Caveats

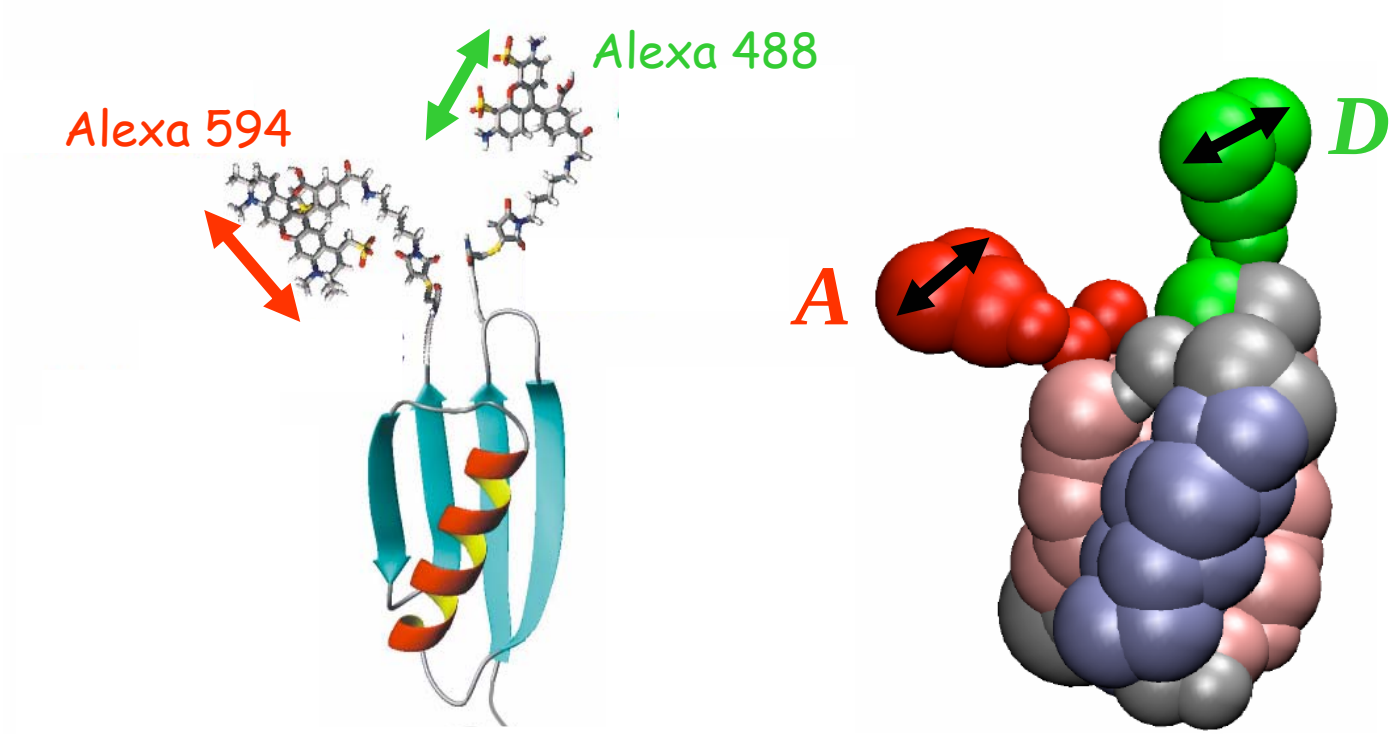
Oversimplified model



FRET may only monitor part of the transition path

Robert Best (U. Cambridge) Langevin simulations of (Brooks, native interactions only) bead model

Calculated $Q(t)$, $R_g(t)$, $r_{DA}(t)$, $\boldsymbol{E(t)}$

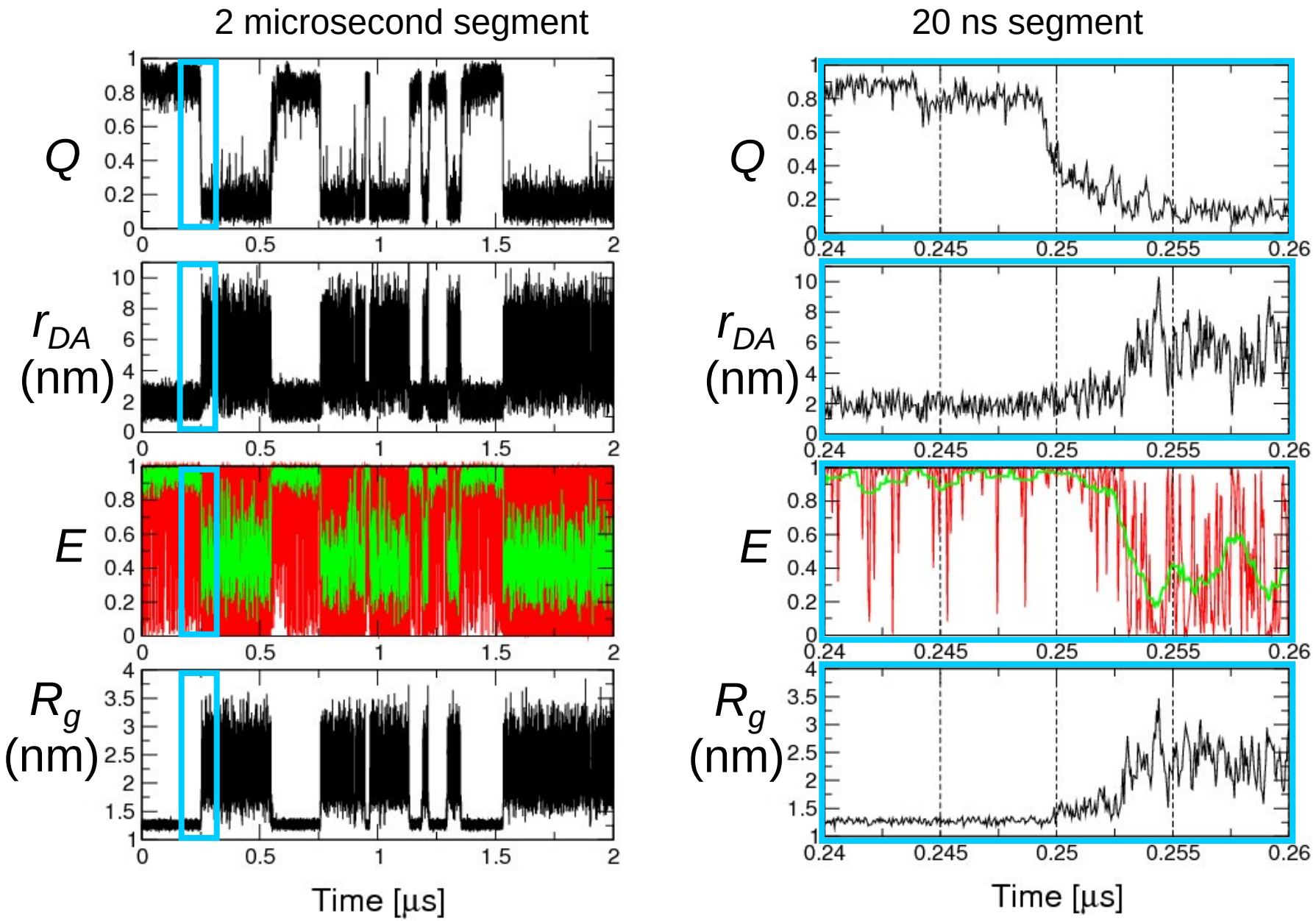


$$\kappa(t) = \hat{\mu}_D(t)\cdot\hat{\mu}_A(t) - 3(\hat{r}_{DA}(t)\cdot\hat{\mu}_D(t))(\hat{r}_{DA}(t)\cdot\hat{\mu}_A(t))$$

$$k_T(t) = k_D R_0^6 \frac{3(\kappa(t))^2}{2(r_{DA}(t))^6}$$

$$E(t) = \frac{k_T(t)}{k_T(t) + k_D}$$

Simulation at 1/500th water friction



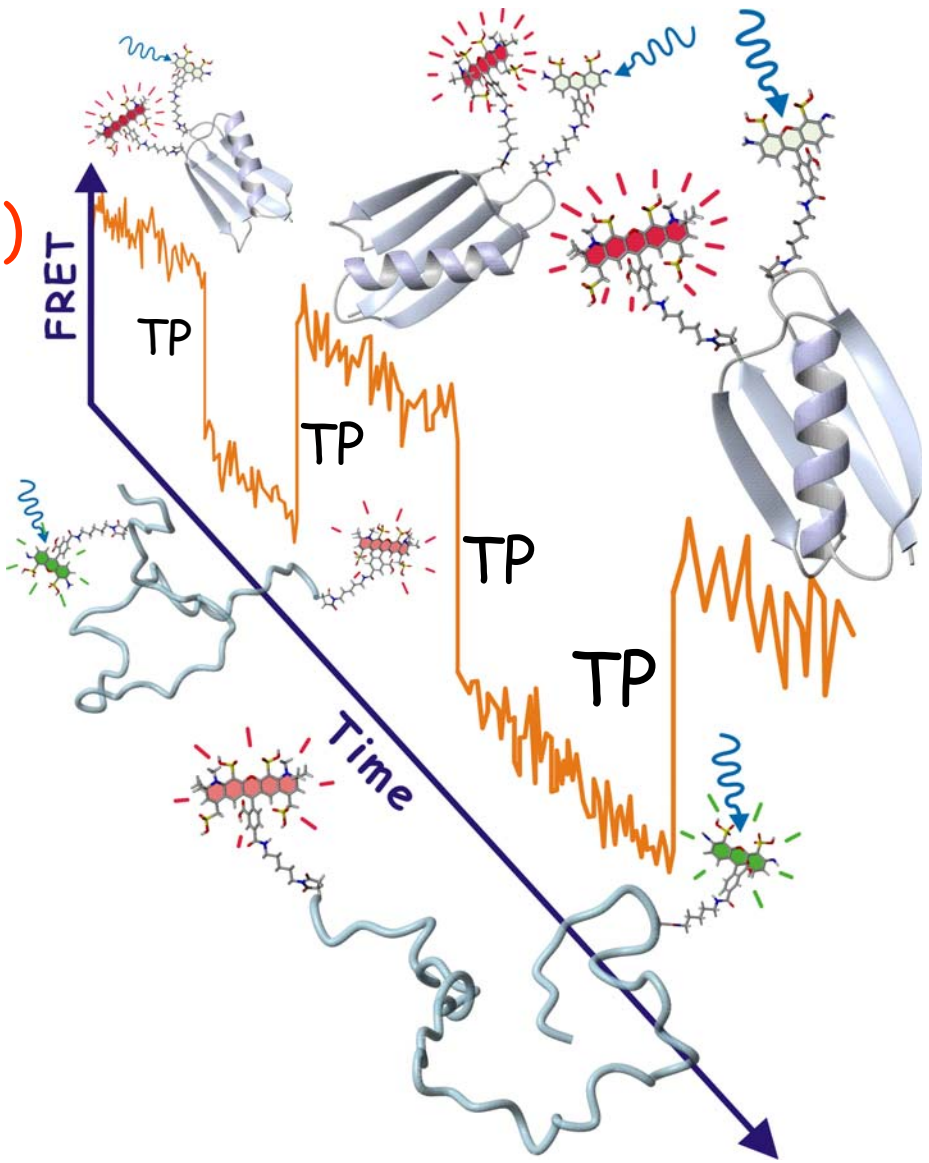
Future Directions

Measure transition path time for WW domain and protein G (?)

Distance versus time measurements during transition path for WW domain (increase viscosity)

Label with > 2 fluorophores: multiple simultaneous distances versus time places major constraints on possible folding mechanisms

Looking forward to lots of interesting results !!



Coworkers in Laboratory of Chemical Physics, NIH, Bethesda

Hoi Sung Chung

Theory of single molecule expts.
Attila Szabo and Irina Gopich

Protein engineering
John M. Louis

Automation
Kevin McHale

Langevin simulations
Robert Best, U. Cambridge