

SUPPLEMENTARY INFORMATION

Mapping fast protein folding with multiple-site fluorescent probes

Maxim B. Prigozhin^{†,^}, Shu-Han Chao^{§,^}, Shahar Sukenik[†], Taras V. Pogorelov^{†,*,^}, Martin Gruebele^{†,§,¶,*}

[†]Departments of Chemistry and [§]Physics, [¶]Center for Biophysics and Computational Biology, and ^{*}National Center for Supercomputing Applications, University of Illinois, Urbana, Illinois 61801, United States

Protein sequences and probe check by molecular dynamics simulation

Protein sequences

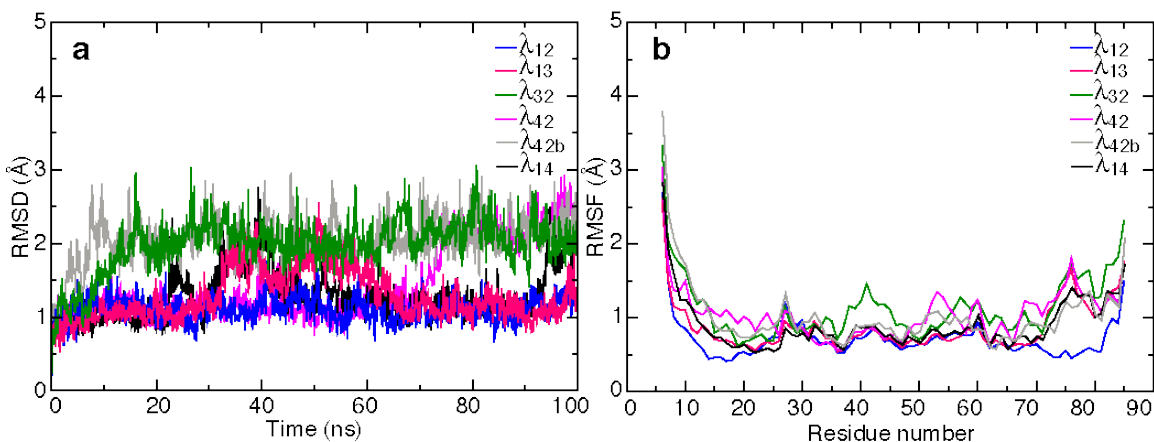
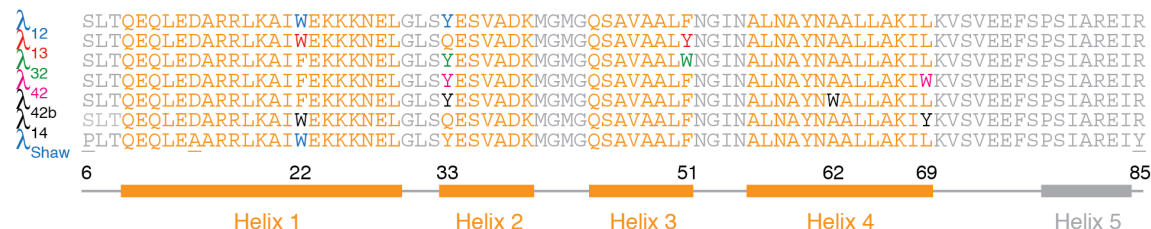


Figure S1. Top: Sequences of seven mutants of λ_{6-85} used in this work including the mutant that was simulated by D. E. Shaw Research (listed last). Three mutations of this simulated sequence with respect to the experimental template sequence λ_{12} are underlined: S6P, D14A, R85Y. Note that all experimental sequences also included the following fragment at the beginning, which was part of the pET-15b vector and was required for protein purification: MGSSHHHHH HSSGLVPRGSHM. Bottom: Structural characteristics of the lambda repressor and its 5 mutants during equilibrium 100 ns MD simulations. (a) Root-mean square deviation (RMSD), based on non-hydrogen atoms of the backbone. (b) Root-mean square fluctuations (RMSF), based on C_{α} atoms. RMSD and RMSF show relative stability of the proteins.

Equilibrium measurements

Temperature melts of λ mutants monitored by circular dichroism spectroscopy

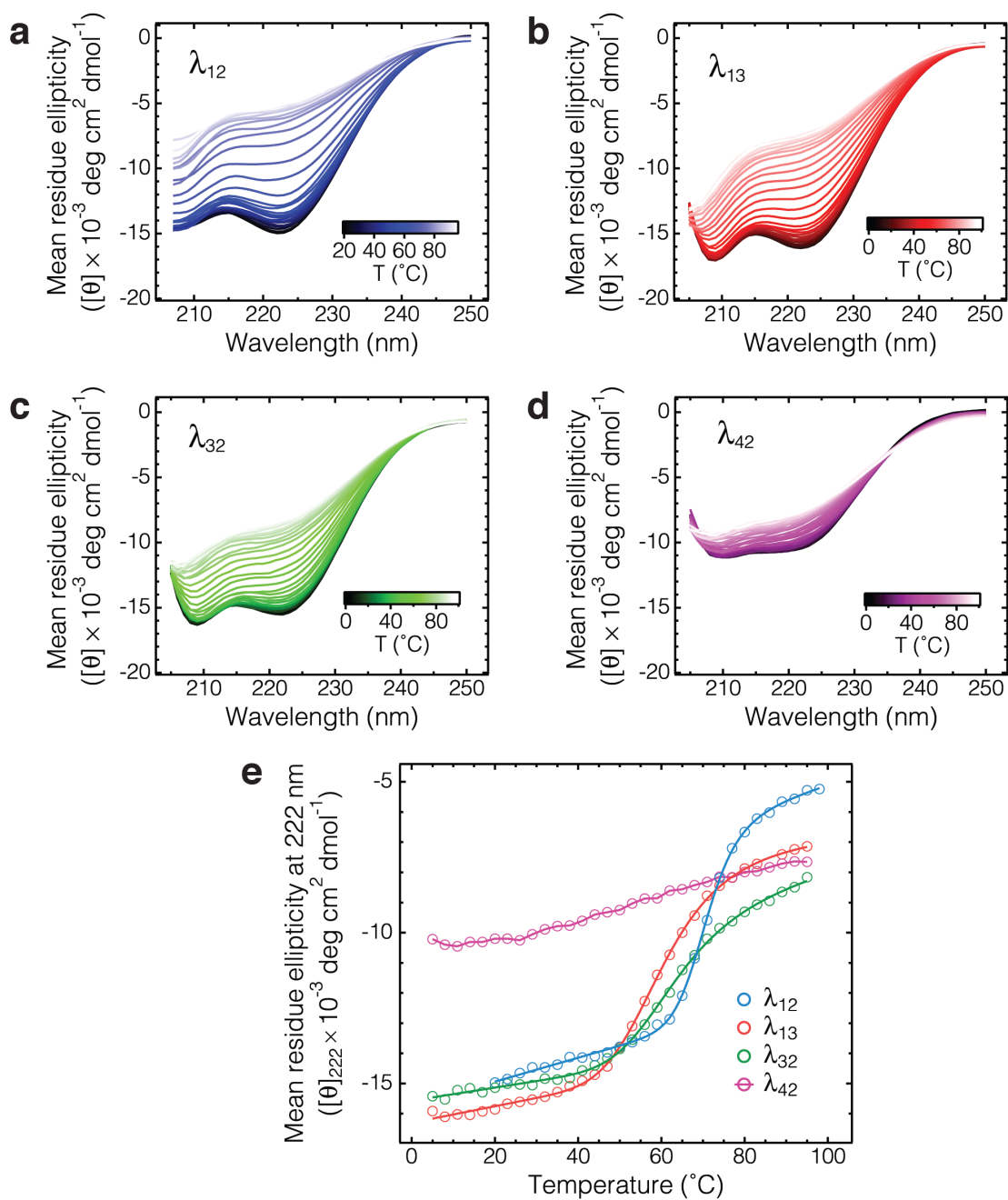


Figure S2. Temperature melts of the new lambda repressor mutants probed by circular dichroism spectroscopy. (a–d) Raw data for λ_{12} , λ_{13} , λ_{32} , and λ_{42} , respectively. The spectra were taken in 3 °C increments. (e) Mean residue ellipticity at 222 nm as a function of temperature for the new mutants. Solid lines are two-state thermodynamic fits. Note that λ_{12} , λ_{13} , and λ_{32} show a cooperative transition upon unfolding, while λ_{42} does not. The profile of λ_{42} was not fitted.

Temperature melts of λ mutants monitored by fluorescence spectroscopy

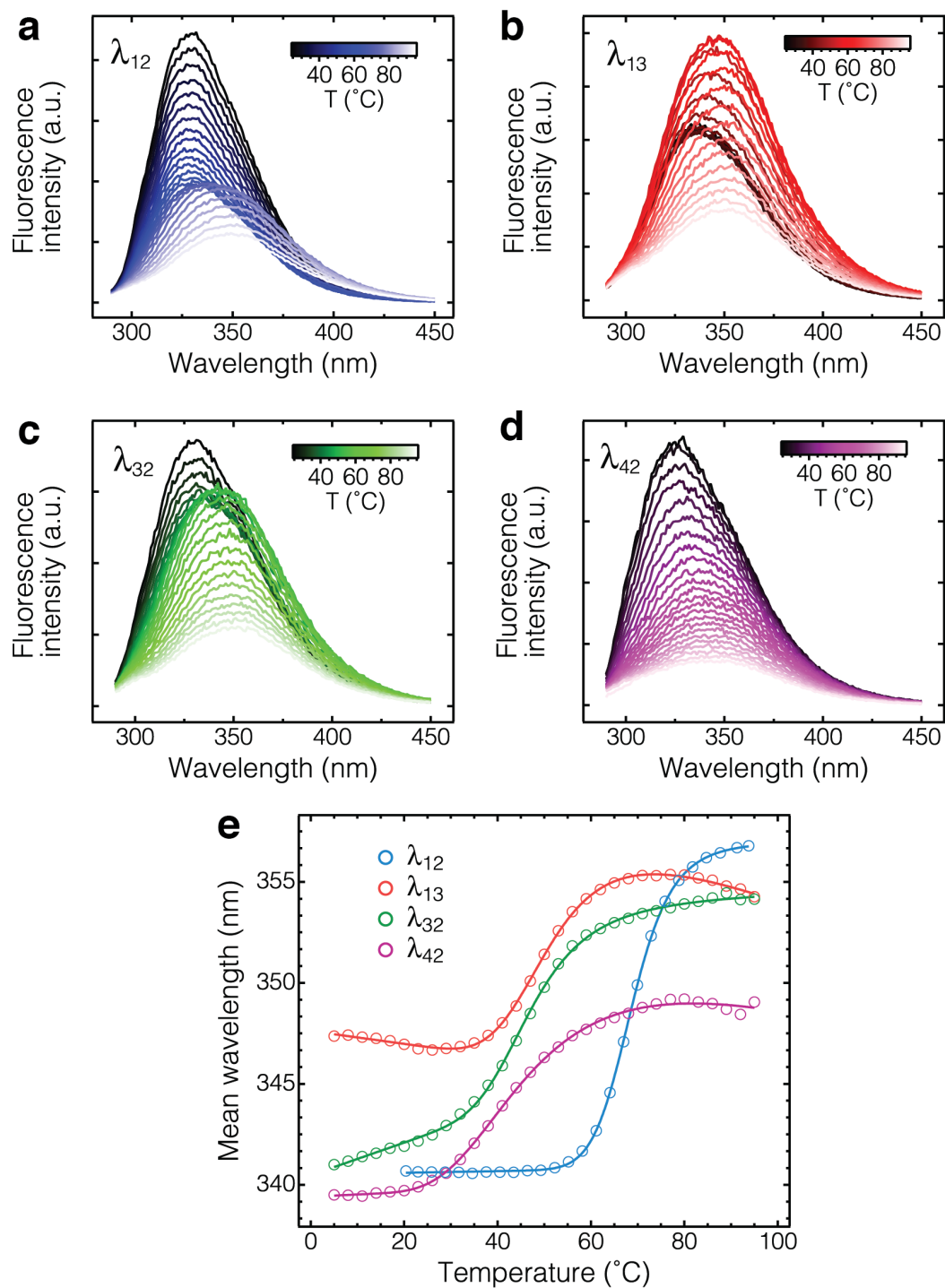


Figure S3. Temperature melts of the new lambda repressor mutants probed by fluorescence spectroscopy. (a–d) Raw data for λ_{12} , λ_{13} , λ_{32} , and λ_{42} , respectively. The spectra were taken in 3 °C increments. (e) Spectral mean of fluorescence spectra of the new lambda mutants as a function of temperature. Solid lines are two-state thermodynamic fits.

Temperature melts of λ_{12} in 2.2 M guanidine hydrochloride (i.e. sample λ_{12}^*)

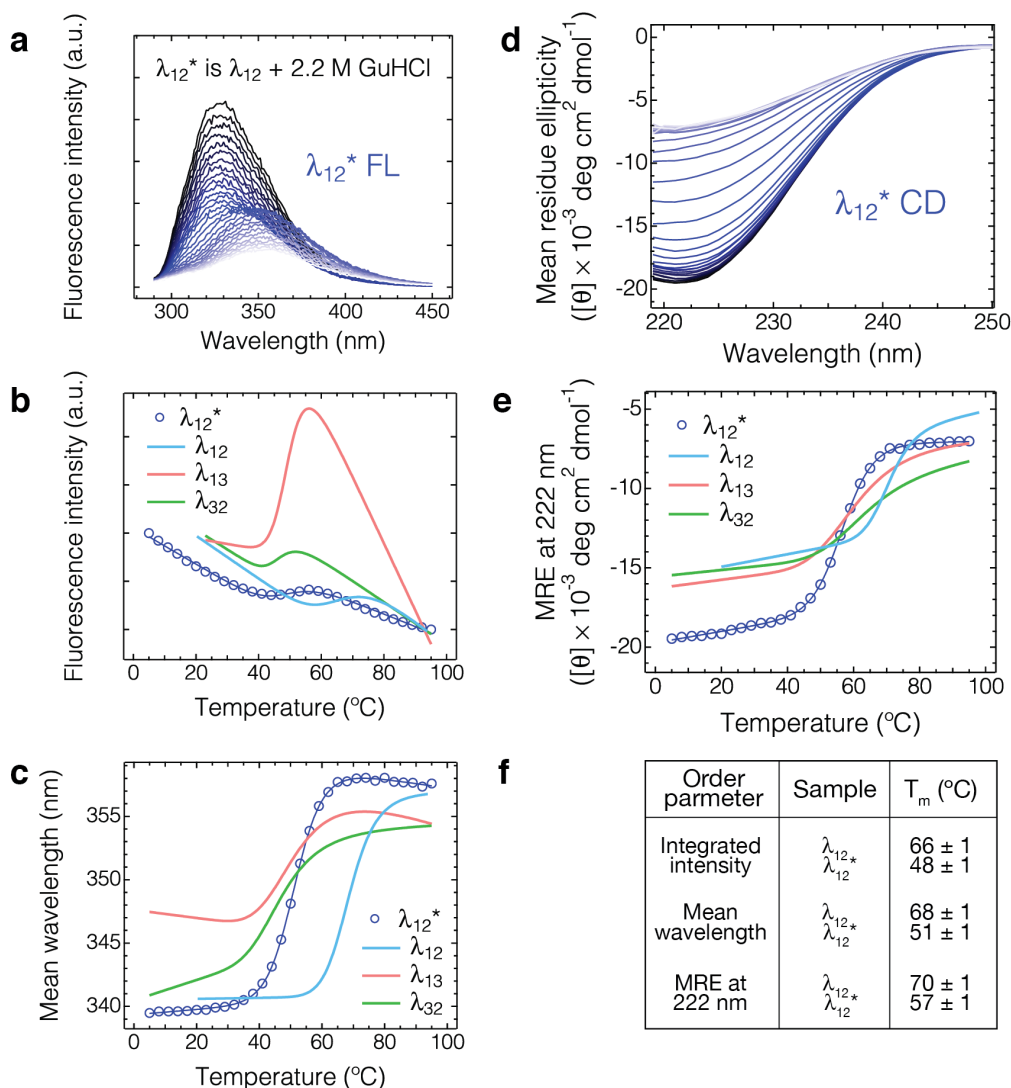


Figure S4. Equilibrium thermal denaturation of λ_{12} with 2.2 M GuHCl (λ_{12}^*). (a) Temperature melt of λ_{12}^* monitored by fluorescence spectroscopy. Spectra range from 5°C (black) to 95°C (light blue) in 3°C increments. (b) Analysis of data in panel (a) in terms of the integrated fluorescence intensity. The data were normalized to start at 1 and end at 0. Solid lines are two-state thermodynamic fits for λ_{12}^* and also λ_{12} , λ_{13} , and λ_{32} for comparison. (c) Analysis of data in panel (a) in terms of the mean wavelength (see Methods for details). Solid lines are two-state thermodynamic fits for λ_{12}^* and also λ_{12} , λ_{13} , and λ_{32} for comparison. (d) Temperature melt of λ_{12}^* monitored by circular dichroism spectropolarimetry. Spectra range from 5°C (black) to 95°C (light blue) in 3°C increments. Absorption of GuHCl prevents measurement at wavelengths below 220 nm. (e) Analysis of data in panel (d) in terms of the mean residue ellipticity at 222 nm as a proxy for α -helical content. Solid lines are two-state thermodynamic fits for λ_{12}^* and also λ_{12} , λ_{13} , and λ_{32} for comparison. Note that the native state baseline is lower for λ_{12}^* than for λ_{12} . This result is expected and has been reported previously for lambda repressor.¹ (f) Tabulated denaturation midpoint temperatures for λ_{12} and λ_{12}^* as monitored by different order parameters. Note the expected destabilization of λ_{12}^* with respect to λ_{12} .

Thermodynamic two-state fits using various probes and order parameters

Table S1. Two-state thermodynamic fitting parameters for the equilibrium temperature melts of the studied mutants. Fitting coefficients for three order parameters are shown: integrated fluorescence intensity, spectral mean of fluorescence band, and mean residue ellipticity at 222 nm for the circular dichroism measurement.

Order parameter	Mutants	Fitting parameters		
		T _m (°C)	ΔH (J mol ⁻¹)	ΔS (J mol ⁻¹ K ⁻¹)
Integrated Intensity	λ ₁₂	66 ± 1	-163765 ± 18813	483 ± 55
	λ ₁₂ *	48 ± 1	-171790 ± 25665	535 ± 80
	λ ₁₃	50 ± 1	-282433 ± 32487	874 ± 99
	λ ₃₂	46 ± 1	-256597 ± 41547	804 ± 129
	λ ₄₂	39 ± 1	-106443 ± 5967	341 ± 17
Mean wavelength	λ ₁₂	68 ± 1	-241534 ± 4100	708 ± 6
	λ ₁₂ *	51 ± 1	-196385 ± 3167	605 ± 10
	λ ₁₃	50 ± 1	-138631 ± 6154	429 ± 17
	λ ₃₂	46 ± 1	-144575 ± 6830	453 ± 19
	λ ₄₂	41 ± 1	-86077 ± 7826	274 ± 24
Mean residue ellipticity at 222 nm	λ ₁₂	70 ± 1	-268343 ± 12281	782 ± 34
	λ ₁₂ *	57 ± 1	-196439 ± 4359	595 ± 13
	λ ₁₃	58 ± 1	-139414 ± 7358	421 ± 21
	λ ₃₂	60 ± 1	-142921 ± 11249	429 ± 33
	λ ₄₂	—	—	—

Kinetic measurements

Dependence of kinetics on protein concentration

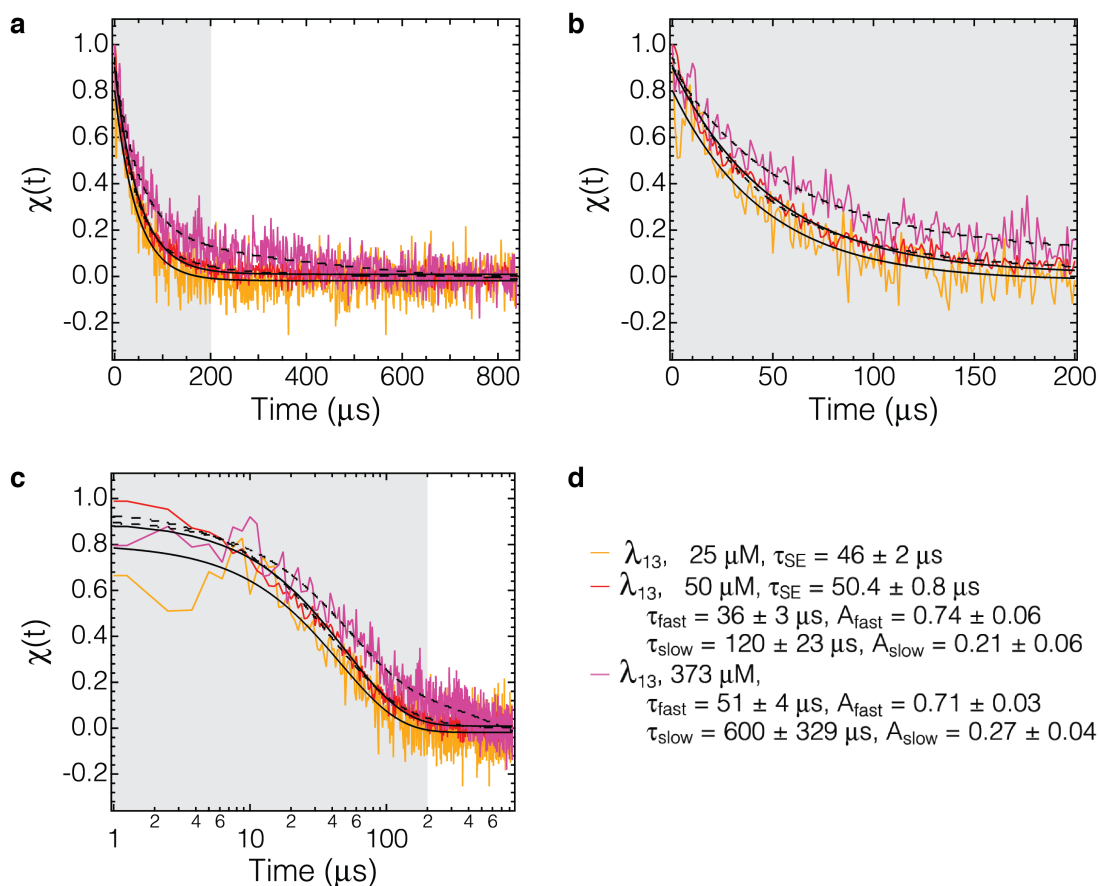


Figure S5. Concentration dependence of λ_{13} kinetics. (a) Kinetic traces of λ_{13} were observed at a final temperature of 54 °C and protein concentrations of 25 μM (yellow), 50 μM (red), and 373 μM (magenta). At higher protein concentration, slight aggregation was observed, which led to the appearance of a slow kinetic phase. Data taken at 25 μM could only be fitted to a single-exponential function (solid black line). Data at 50 μM could be fitted to both a single-exponential function (solid black line) and a double-exponential function (dashed black line). At 373 μM , the data could only be fitted to a double-exponential function (dashed black line). (b) A zoom in of the first 200 μs of the kinetic traces (gray background in all panels). (c) The same data as in panel (a) plotted with logarithmic x -axis to highlight the fast phase. (d) Fitting parameters used for the fits in panels (a), (b), and (c).

Kinetic relaxation at several temperatures

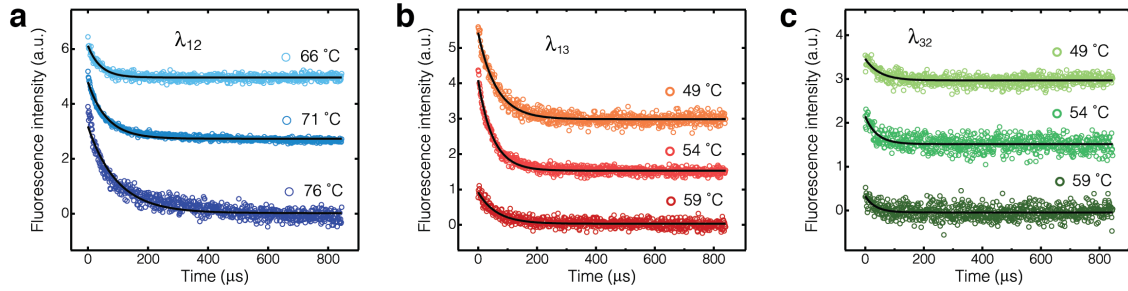


Figure S6. Temperature dependence of kinetic relaxation traces collected at 50 μM protein concentration. Kinetic traces of lambda repressor mutants λ_{12} , λ_{13} , and λ_{32} were observed after a 9 $^{\circ}\text{C}$ temperature jump to the final temperatures indicated in panels (a), (b), and (c), respectively.

Table S2. Time constants for the kinetic relaxation traces of the mutants studied at 50 μM concentration from Figure S6. Single-exponential relaxation time constants were fitted using an equation $FI(t) = A \exp(-t/\tau)$.

Mutant	Temperature	τ (μs)
λ_{12}	66 $^{\circ}\text{C}$	43 ± 2
	71 $^{\circ}\text{C}$	63 ± 1
	76 $^{\circ}\text{C}$	97 ± 3
λ_{13}	49 $^{\circ}\text{C}$	62 ± 1
	54 $^{\circ}\text{C}$	50.4 ± 0.8
	59 $^{\circ}\text{C}$	63 ± 3
λ_{32}	49 $^{\circ}\text{C}$	49 ± 5
	54 $^{\circ}\text{C}$	39 ± 5
	59 $^{\circ}\text{C}$	35 ± 7

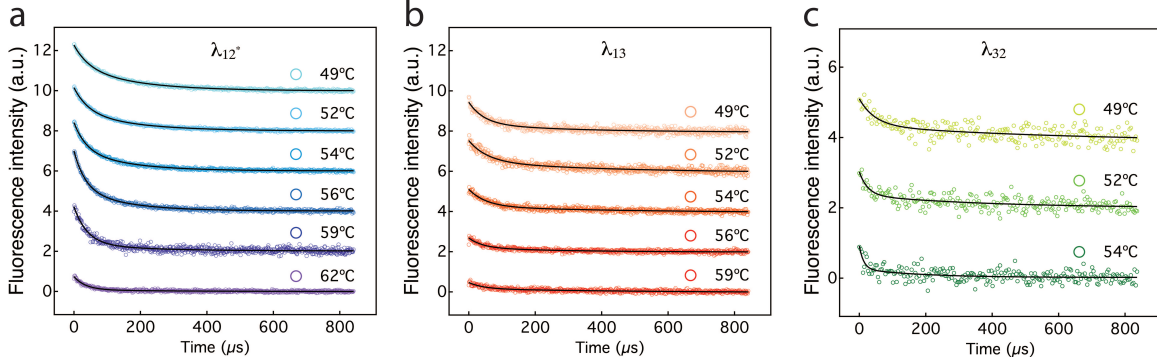


Figure S7. Temperature dependence of kinetic relaxation traces collected at the concentrations of 309 μM , 373 μM , and 140 μM for λ_{12} with 2.2 M GuHCl (λ_{12}^*), λ_{13} , and λ_{32} , respectively. Kinetic traces of lambda repressor mutants λ_{12}^* , λ_{13} , and λ_{32} were observed after a 8 °C temperature jump to the final temperatures indicated in panels (a), (b), and (c), respectively. Double-exponential fits are shown as solid black lines. The fitting parameters are listed in Table S3. Data of λ_{32} has a three times larger binning period than that of λ_{12}^* and λ_{13} .

Table S3. Fitting parameters for the kinetic relaxation traces of the data shown in Figure S7. Double-exponential relaxation time constants were fitted using an equation $FI(t) = A_1 \exp(-t/\tau_1) + A_2 \exp(-t/\tau_2)$. The protein folding time constant is highlighted in bold font.

Temperature	Mutants	Fitting parameters			
		A_1	τ_1 (μs)	A_2	τ_2 (μs)
49°C	λ_{12}^*	1.06 ± 0.06	45 ± 3	1.21 ± 0.06	177 ± 7
	λ_{13}	1.07 ± 0.08	45 ± 6	0.42 ± 0.07	193 ± 93
	λ_{32}	0.75 ± 0.16	47 ± 19	0.47 ± 0.22	625 ± 1020
52°C	λ_{12}^*	1.17 ± 0.05	43 ± 2	0.99 ± 0.05	174 ± 7
	λ_{13}	1.04 ± 0.09	56 ± 7	0.56 ± 0.05	415 ± 158
	λ_{32}	0.66 ± 0.12	26 ± 10	0.37 ± 0.06	381 ± 277
54°C	λ_{12}^*	1.32 ± 0.06	36 ± 2	1.09 ± 0.06	149 ± 6
	λ_{13}	0.87 ± 0.04	53 ± 5	0.32 ± 0.03	548 ± 291
	λ_{32}	0.65 ± 0.14	13 ± 6	0.25 ± 0.08	186 ± 104
56°C	λ_{12}^*	1.83 ± 0.07	34 ± 2	1.16 ± 0.07	141 ± 7
	λ_{13}	0.48 ± 0.04	41 ± 6	0.22 ± 0.03	288 ± 86
	λ_{32}	–	–	–	–
59°C	λ_{12}^*	1.83 ± 0.11	39 ± 3	0.37 ± 0.11	197 ± 63
	λ_{13}	0.31 ± 0.03	44 ± 8	0.21 ± 0.04	620 ± 398
	λ_{32}	–	–	–	–
61°C	λ_{12}^*	0.66 ± 0.02	35 ± 2	0.09 ± 0.01	307 ± 131
	λ_{13}	–	–	–	–
	λ_{32}	–	–	–	–

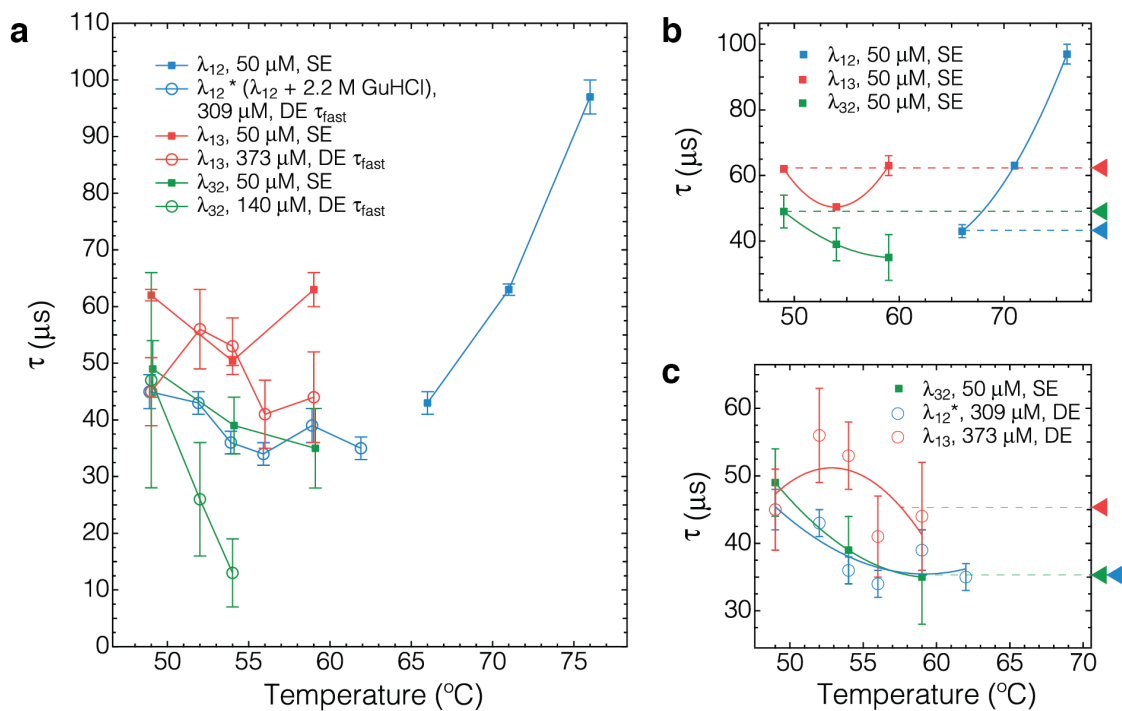


Figure S8. Kinetics of lambda repressor multi-site probe mutants. (a) All the data for a series of mutants collected at various protein and guanidine hydrochloride (GuHCl) concentrations. The data were collected for different protein expressions and by different researchers to provide a picture of the consistency that can be obtained under different conditions. SE means that the data was fitted using a single-exponential fit (data in Figure S6, fit parameters in Table S2). DE τ_{fast} means that the data was fitted using a double-exponential fit and the fast phase is shown (data in Figure S7, fit parameters in Table S3, fast phase highlighted in bold font). λ_{12} is in blue, λ_{13} is in red, and λ_{32} is in green. λ_{12}^* corresponds to λ_{12} with 2.2 M GuHCl. Squares represent data taken at 50 μM protein concentration. Circles represent data taken at 309 μM , 373 μM , and 140 μM for λ_{12} with 2.2 M GuHCl (λ_{12}^*), λ_{13} , and λ_{32} , respectively. Error bars represent the fitting errors. (b) A plot of data for only the measurements collected at 50 μM protein concentration and without GuHCl. λ_{12} is more stable than λ_{13} and λ_{32} . However, if time constants at temperatures that correspond to the melting midpoint temperatures of each respective protein are compared, the ordering of contact formation rate is $\lambda_{12} = \lambda_{32} > \lambda_{13}$ going from the fastest to the slowest. λ_{12} and λ_{32} are indistinguishable within the experimental uncertainty. Solid arrows on the right highlight the observed trend. Solid parabolas are guides for the eye. (c) A plot of data for measurements where more temperature points were collected for λ_{12} and λ_{13} albeit at a higher protein concentration. 2.2 M GuHCl was used to destabilize λ_{12} and compare it directly with the other two mutants at the same temperatures. The order of contact formation is still $\lambda_{12} = \lambda_{32} > \lambda_{13}$ going from the fastest to the slowest. Solid arrows on the right highlight the observed trend. Solid parabolas are guides for the eye.

MD simulation analysis

Analysis of MD simulations

Data from D. E. Shaw Research² were recorded every 200 ps, yielding 805,107 frames. All frames were used for our analysis. MD data generated by the authors (the 100 ns simulations) were collected at 100 ps intervals.

Table S4. Distance between fluorophore side chains and solvent accessible surface area (SASA). SASA was measured with a probe of 1.4 Å radius in VMD. Mean values and standard deviations were calculated using 1,000 data points from 100 ns long trajectories. For comparison, SASA of the completely exposed side chains are 279.5 Å (W) and 253 Å (Y).

	λ_{12}	λ_{13}	λ_{32}	λ_{14}	λ_{42}	λ_{42b}
W/Y position	22/33	22/51	51/33	22/69	69/33	62/33
Distance (Å)	5.8±0.4	4.4±0.5	6.2±0.5	6.0±0.6	10.1±1.0	14.5±0.8
W SASA (Å ²)	8.4±7.4	15.0±9.1	40.7±19.3	17.1±10.2	5.7±3.9	41.9±14.6
Y SASA (Å ²)	75.3±11.4	9.0±6.0	60.3±16.1	1.8±3.27	62.5±21.1	62.1±24.3

Order parameters

The distance between residues was defined to be the distance between centers of mass of the side chains. This distance, as well as RMSF and RMSD were calculated using the “measure” script in VMD.

The fraction of native contacts (Q) is the order parameter that measures the similarity of a given structure to a structure in the native state as defined based on Q by Wolynes *et al.*³ We used the alpha carbon atom distances for atoms with $|i-j| \geq 7$ excluding nearest neighbors:

$$Q(t) = \frac{2}{(\#C_{\alpha} - 6)(\#C_{\alpha} - 7)} \sum_{i=1}^{\#C_{\alpha}} \sum_{\substack{j=1 \\ |j-i| > 6}}^{\#C_{\alpha}} \exp \left[- (d_{ij}(t) - d_{ij}^{Nat})^2 / 2|i-j|^2 \right]$$

Distributions of the inter-helical distances were calculated with the “histe” command of Matlab with the distance data binned into 0.5 Å intervals. The free energy in the order parameters of choice was calculated as $FE = -k_B T \ln P$, where P is the probability of observing a particular value of the order parameters based on the distribution, k_B – Boltzmann constant, and T – temperature.

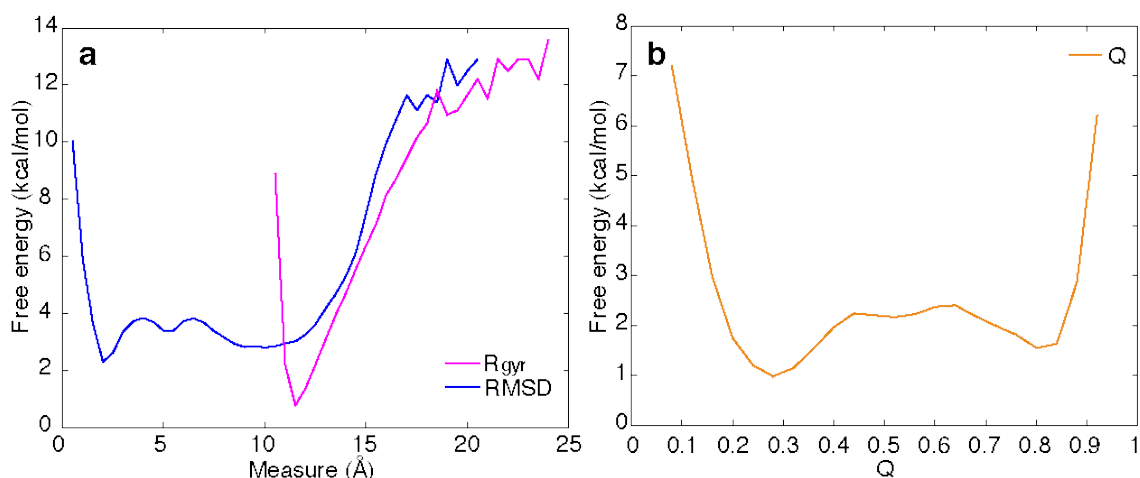


Figure S9. Free energy profiles of lambda repressor along various order parameters (161 μ s simulation data set courtesy of D. E. Shaw Research): (a) RMSD and R_{gyr} and (b) the fraction of native contacts, Q .

Autocorrelation functions

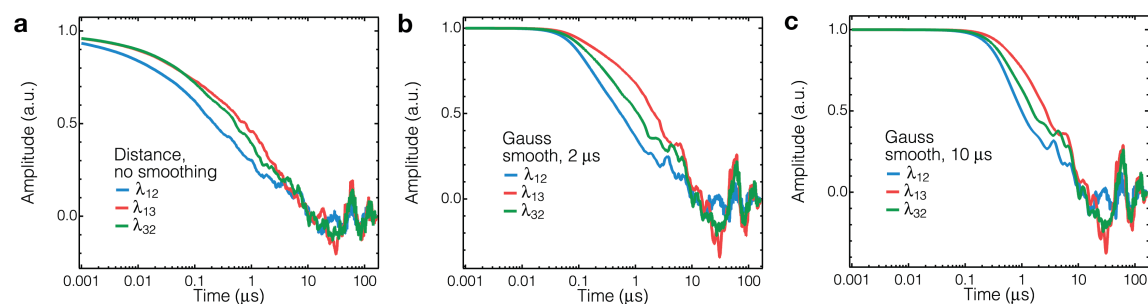


Figure S10. The autocorrelation in Fig. 3b most closely mimics the experimental observable (Dexter-quenched Trp-Tyr with a quenching efficiency $\delta_{ij} \sim \exp(-\Delta_{ij}/a)$, with fitted double exponential times λ_{13} : 160 ± 10 ns and 4.48 ± 0.01 μ s; λ_{32} : 191 ± 5 ns and 4.26 ± 0.01 μ s; λ_{12} : 262 ± 2 ns and 4.22 ± 0.01 μ s). However, autocorrelation analysis of molecular dynamics time trajectories of distances Δ_{ij} without Dexter transfer efficiency scaling produce the same time ordering. For Fig. 3b in the main text and Fig. S10 above, Matlab function *xcov* was used to obtain the autocorrelation of the time trajectory. The ordering is not sensitive to time-smoothing of the MD trajectory, as can be seen in panels a-c.

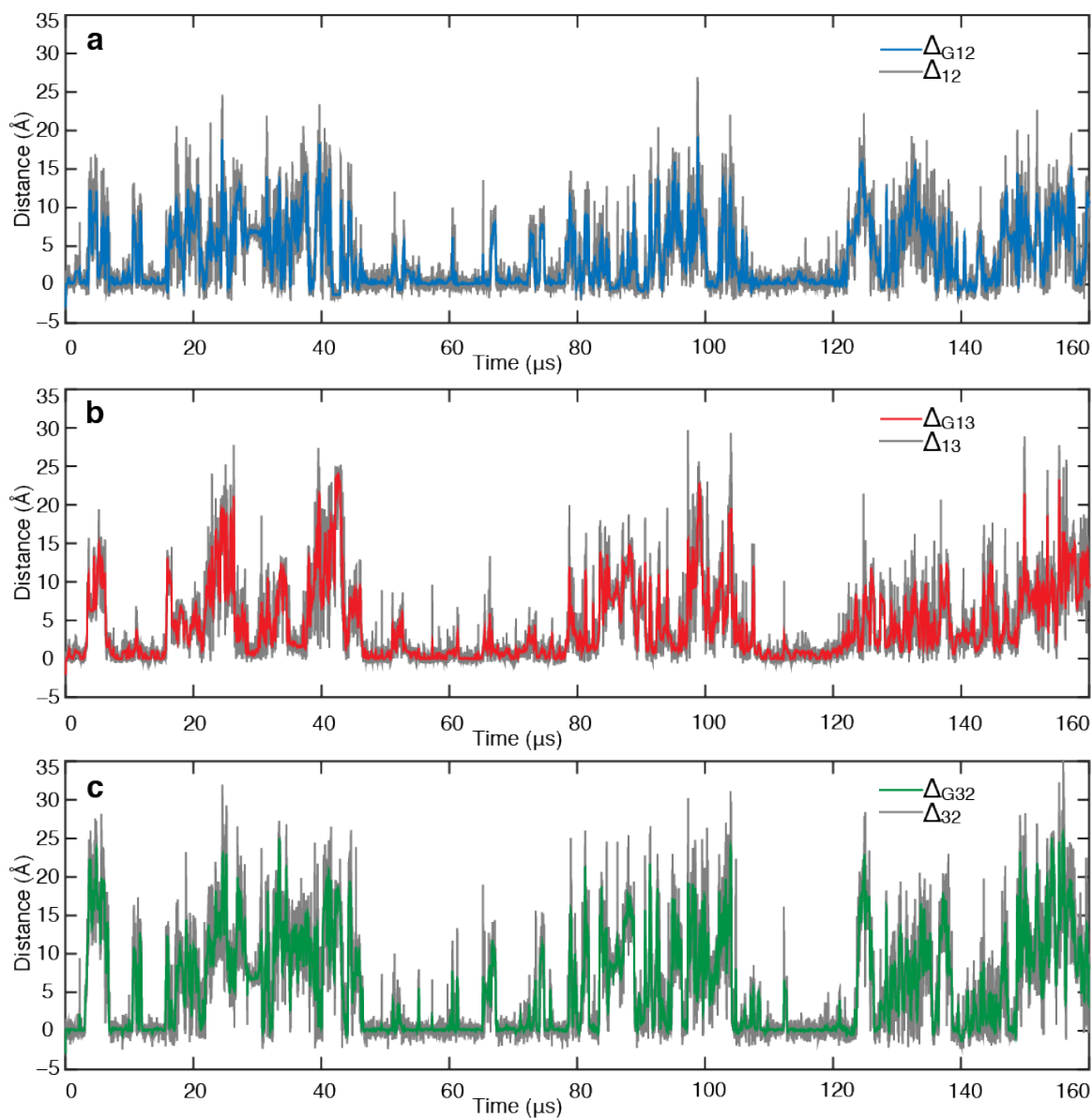


Figure S11. Smoothed (by a 2 μs Gaussian filter) time traces of distances (with respect to the native state values) between side chains of the simulated mutant where Trp and Tyr are located in λ_{12} , λ_{13} , and λ_{32} (shown in color) and raw time traces (shown in gray).

Transition matrix

To calculate the transition matrix in Fig. 4j, we defined three states: native (N) with $Q > 0.6$ and short inter-probe distances, trapped state (T) where $Q > 0.6$, the distance between helices e.g. 1 and 3 (Δ_{13}) is native-like (short) but distances between helices 1 and 2 (Δ_{12}) and helices 3 and 2 (Δ_{32}) are longer than in a native structure ($\Delta_{13} < 3 \text{ \AA}$ while $\Delta_{12} > 5 \text{ \AA}$ and $\Delta_{32} > 5 \text{ \AA}$), and denatured (D) state, which included all other possibilities. Transitions between and within the three states for three possible models were counted and probability of transition was normalized by the state. The matrix calculations were performed in Matlab.

Table S5. Markov model is presented for the three-state system used for analysis of the off-pathway trap state with $\Delta_{13} < 3 \text{ \AA}$ while $\Delta_{12} > 5 \text{ \AA}$ and $\Delta_{32} > 5 \text{ \AA}$ (161 μs simulation data set courtesy of D. E. Shaw Research).

State	Native (N)	Trapped (T) $\Delta_{13} < 3 \text{ \AA}$	Denatured (D)
Native (N)	210,714	436	4,179
Trapped (T) $\Delta_{13} < 3 \text{ \AA}$	438	6,272	310
Denatured (D)	4,176	312	578,269

Table S6. Markov model is presented for the three-state system used for analysis of the potential off-pathway trap state with $\Delta_{12} < 3 \text{ \AA}$ while $\Delta_{13} > 5 \text{ \AA}$ and $\Delta_{32} > 5 \text{ \AA}$ (161 μs simulation data set courtesy of D. E. Shaw Research).

State	Native (N)	Trapped (T) $\Delta_{12} < 3 \text{ \AA}$	Denatured (D)
Native (N)	217,766	9	4,454
Trapped (T) $\Delta_{12} < 3 \text{ \AA}$	7	78	35
Denatured (D)	4,455	33	578,269

Table S7. Markov model is presented for the three-state system used for analysis of the potential off-pathway trap state with $\Delta_{32} < 3 \text{ \AA}$ while $\Delta_{12} > 5 \text{ \AA}$ and $\Delta_{13} > 5 \text{ \AA}$ (161 μs simulation data set courtesy of D. E. Shaw Research).

State	Native (N)	Trapped (T) $\Delta_{32} < 3 \text{ \AA}$	Denatured (D)
Native (N)	217,834	11	4,489
Trapped (T) $\Delta_{32} < 3 \text{ \AA}$	11	4	0
Denatured (D)	4,455	0	578,269

Mean first passage time

Mean first passage times on the estimated free energy profiles (Fig. 4b) were calculated using Matlab as follows:⁴

$$MFPT = \int_{r_1}^{r_2} dy \frac{\exp(\beta FE(y))}{D(y)} \int_r^y \exp(-\beta FE(z)) dz$$

where $FE(y)$ is the free energy estimate, $D(y)$ is diffusion coefficient as a function of order parameter y , and $\beta = 1/k_B T$.

The free energy along the order parameter of choice (y) was estimated as $FE(y) = -k_B T \ln\left(\frac{P_i}{\Delta y}\right)$, where P_i is the probability of observing a value of the order parameter in the bin i based on the distribution of the MD data.

Diffusion coefficients $D(y)$ were estimated as follows:⁵

$$D(y)_{i+\frac{1}{2}} \approx \Delta y^2 R_{i+1,i} \left(\frac{P_i}{P_{i+1}}\right)^{1/2},$$

where i and $i+1$ are neighboring bins of y_i and y_{i+1} and R_{ij} – elements of the rate matrix that for short Δt were approximated as follows:

$$R_{ij} \Delta t \approx \frac{N_{ij} + N_{ji}}{N_{ii} P_j + N_{jj} P_i} P_i,$$

where N_{ij} – number of transitions from bin j to bin i .

Table S8. Mean first passage time for Dexter-scaled MD trajectories binned into 20 bins for varied lag time (Δt , ns) as compared to the time scales from experiments and MD autocorrelation functions.

	Experiment τ , μs	Computation autocorrelation τ , μs	MFPT, μs					
			$\Delta t=0.2$	$\Delta t=1$	$\Delta t=2$	$\Delta t=5$	$\Delta t=10$	$\Delta t=20$
λ_{12}	43	4.22	0.038	0.18	0.37	0.91	1.85	4.2
λ_{13}	62	4.48	0.063	0.30	0.60	1.39	2.81	5.2
λ_{32}	49	4.26	0.047	0.21	0.42	1.15	1.88	4.0

References:

1. Gelman, H., Perlova, T. & Gruebele, M. Dodine as a protein denaturant: the best of two worlds? *J. Phys. Chem. B* **117**, 13090–7 (2013).
2. Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How fast-folding proteins fold. *Science* **334**, 517–20 (2011).
3. Eastwood, M. P., Hardin, C., Luthey-Schulten, Z. & Wolynes, P. G. Evaluating protein structure-prediction schemes using energy landscape theory. *IBM J. Res. Dev.* **45**, 475–97 (2001).
4. Schulten, K., Schulten, Z. & Szabo, A. Dynamics of reactions involving diffusive barrier crossing. *J. Chem. Phys.* **74**, 4426–32 (1981).
5. Hummer, G. Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New J. Phys.* **7**, 34 (2005).