LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-208639

Probing structural heterogeneities and conformational fluctuations of biopolymers

T. Laurence, X. Kong, M. Jaeger, S. Weiss

December 16, 2004

Proceedings of the National Academy of Sciences of the United States of America

Probing structural heterogeneities and conformational fluctuations of biopolymers

Ted A. Laurence[A,*,#], Xiangxu Kong[B,*], Marcus Jaeger[B], Shimon Weiss[B, C,#]

[A]*Physical Biosciences Institute, Lawrence Livermore National Laboratory, Livermore, California, 94550*

[B]*Department of Chemistry and Biochemistry, and*

[C]*Department of Physiology, University of California Los Angeles, Los Angeles, California, 90095*

**These authors contributed equally to the work*

#Correspondence should be addressed to T.L. and S.W. (e-mail: laurence2@llnl.gov, sweiss@chem.ucla.edu).

We study protein and nucleic acid structure and dynamics using single-molecule fluorescence resonance energy transfer measurements with alternating-laser excitation. Freely diffusing molecules are sorted into subpopulations based on stoichiometry, detecting donor and acceptor coincidence for periods over 100 µs-1 ms. Faster (< 100 µs) fluctuating distance distributions are studied within these subpopulations using time-resolved single photon counting measurements. We find that short double-stranded DNA (dsDNA) is more flexible than expected from persistence lengths measured on long dsDNA. We find that the electrostatic portion of the persistence length of single-stranded poly-dT varies as the ionic strength (I) to the -½ power ($I^{-1/2}$). Lastly, we find that the unfolded protein Chymotrypsin Inhibitor 2 (CI2) is unstructured at high denaturant. However, in the presence of folded CI2 (at lower denaturant), unfolded CI2 is more compact

and displays larger distance fluctuations, possibly due to unsuccessful attempts to cross the folding barrier.

In contrast to neutral homopolymers[1], biopolymers such as proteins and nucleic acids contain long-range Coulomb interactions[2] and specific, intra-chain interactions[3] that strongly affect their structure and dynamics. Protein folding is the most spectacular manifestation of these interactions[4]. Understanding biopolymer energy landscapes requires measurements of fluctuating distance distributions occurring over short distances (0.1 – 100 nm) and many time scales (picoseconds to minutes). The structure and dynamics of charged polymer chains (polyelectrolytes), including double-stranded DNA (dsDNA) and single-stranded DNA (ssDNA), are strongly affected by long-range, electrostatic repulsion[2]. Since simulations and theory have focused on dilute, single-chain properties, a regime difficult to access by conventional methods, experimental validation of many predictions has been lacking[5]. In protein folding studies, folded, unfolded, and partially folded species may be simultaneously present and rapidly inter-converting, obscuring the properties of individual species. An experimental method that unravels distance distributions and fast conformational fluctuations is therefore in great need.

Fluorescence resonance energy transfer (FRET), the non-radiative transfer of excitation energy from a donor (*D*) to an acceptor (*A*) fluorophore, provides dynamical and structural information on fluctuating biopolymers labeled with *D* and *A*. FRET efficiency *E* depends on the *D*-*A* distance *R*: $E = \left[1 + \left(R/R_0\right)^6\right]$ (at $R=R_0$, $E=0.5$), providing a "spectroscopic ruler" for the 2-8 nm scale[6]. Time-resolved fluorescence lifetime measurements for FRET (TR-FRET) allow analysis of multi-exponential decays related to distance distributions[7]. Previous single-molecule FRET studies[8-10], including those using TR-FRET[11], measured distance distributions with time resolutions of ~100 μs.

Here, we use a novel single molecule TR-FRET spectroscopy, dubbed nanosecond alternating-laser excitation (nsALEX) fluorescence activated molecular sorting (FAMS) that reveals distance distributions within subpopulations of fluorescently labeled biopolymers fluctuating on time scales down to the fluorescence lifetime (~1 ns). We compare the mean and width of the end-to-end distance distributions of 3 types of biopolymers: dsDNA, poly (dT) ssDNA, and the protein Chymotrypsin Inhibitor 2 (CI2). We find that short dsDNA (2-8 nm) is more flexible than expected given the persistence length of ~50 nm measured on long dsDNA. We find that the additional contour length per base for poly (dT) ssDNA is 0.40±0.03 nm, possibly indicating limited base stacking conformers. The electrostatic portion of the persistence length of ssDNA is found to vary as $I^{-1/2}$ (*I* is the ionic strength). Finally, unfolded CI2 (6M GdnCl) is more flexible than ssDNA and behaves as a Gaussian chain. At lower denaturant concentration and in the presence of the folded species, unfolded CI2 adopts more compact states with *larger* distance fluctuations.

"Phase diagram" for polymer fluctuations:

Since $E=\left[1+\left(R/R_0\right)^6\right]$, a polymer's distance distribution $P(R)$ causes a distribution in E ($P(E)$) and hence, multi-exponential fluorescence lifetime decays. $P(R)$ and $P(E)$ for biopolymers are described by contour length L and persistence length l_P [12]. For $L << l_P$, R is fixed (rigid rod; cyan, magenta, and black polymers, figure 1a). For $L >> l_P$ (green polymer, figure 1a), a Gaussian chain describes $P(R)$: $P(R)=4\pi R^2\exp\left(-3R^2/4Ll_P\right)\left(3/4\pi Ll_P\right)^{\frac{3}{2}}$ [12]. For $l_P \approx L$, a wormlike chain model is required. We use an analytic approximation valid over all ratios L/l_P [13]:

$$P(R)=4\pi R^2 N(L,l_P)\exp\left(-3\frac{L}{l_P}\Big/4\left(1-\left(\frac{R}{L}\right)^2\right)\right)\left(1-\left(\frac{R}{L}\right)^2\right)^{-\frac{9}{2}} \qquad (1)$$

$N(L,l_P)=\frac{4(3L/4l_P)^{3/2}\exp(3L/4l_P)}{\pi^{3/2}\left(4+12/(3L/4l_P)+15/(3L/4l_P)^2\right)}$. Other analytic expressions for high[14] or low[15] L/l_P cover smaller ranges in L/l_P.

FRET depends on the relative orientation of the excitation and emission dipoles of *D* and *A* through the factor κ^2 ($R_0 \propto \kappa^2$). For a rigid rod with rapid fluorophore rotational diffusion (cyan polymer, figure 1a), κ^2 is replaced by $\langle\kappa^2\rangle = 2/3$ and *E* is constant. Dipole orientation restrictions and rotational diffusion on time scales near the fluorescence lifetime shift and widen $P(E)$ [16, 17]. Time-resolved fluorescence anisotropy (TR-FA) partially characterizes κ^2 using rotational diffusion timescales τ_r^D and τ_r^A for *D* and *A*. We model κ^2 effects using dipoles with complete orientational freedom, but with incomplete orientational averaging during the fluorescence lifetime[8], an appropriate model for disordered systems and systems using long, flexible fluorophore linkers[17] (magenta polymer, figure 1a; supplement S5).

Fluorophore tethers (typically ≥ 6 carbon bonds) widen distance distributions, which we model as one effective Gaussian chain with the parameter, $r^{\text{tether}} = \sqrt{L^{\text{tether}} l_P^{\text{tether}}}$ (black polymer, figure 1a). We include tether and κ^2 effects in our simulations of biopolymer fluctuations (green polymer, figure 1a).

Figures 1b and 1c show simulated *D* and *A* lifetime decays for the polymers in figure 1a with parameters extracted from experiments. Comparison of the magenta, black, and green lines with the "constant E" decays (cyan lines) show the increasing effects of rotational diffusion, fluorophore tethers, and biopolymer $P(R)$ on the fluorescence lifetime decays.

To extract information on $P(E)$, we approximate multi-exponential fluorescence lifetime decays using a model with two states of constant FRET efficiencies E_1 and E_2,

spending a fraction of time β in state 1 (figure 1d). More sophisticated fitting and deconvolution methods will be pursued in future studies[7, 18-21]. The *D* decay is

$$f^D(\tau) = \beta Q_D / \tau_{f,0}^D \exp\left(-\tau/\tau_{f,1}^D\right) + (1-\beta) Q_D/\tau_{f,0}^D \exp\left(-\tau/\tau_{f,2}^D\right). \quad (2)$$

Q_D is the quantum efficiency. *D* lifetime $\tau_{f,1}^D$ is reduced by FRET from its value $\tau_{f,0}^D$ in the absence of *A*: $\tau_{f,1}^D = (1-E_1)\tau_{f,0}^D$. Likewise, $\tau_{f,2}^D = (1-E_2)\tau_{f,0}^D$. The *A* decay due to FRET is

$$f^E(\tau) = \left[\beta E_1/\tau_{f,1}^D \exp\left(-\tau/\tau_{f,1}^D\right) + (1-\beta) E_2/\tau_{f,2}^D \exp\left(-\tau/\tau_{f,2}^D\right)\right] * f^A(\tau) \quad (3)$$

$*$ denotes convolution. $f^A(\tau) = Q_A/\tau_f^A \exp\left(-\tau/\tau_f^A\right)$ is the intrinsic fluorescence decay of *A*, where τ_f^A and Q_A are defined as for *D*. The FRET rates $E/\tau_{f,1}^D$ and $E/\tau_{f,2}^D$ in equation (3) replace the fluorescence rates $Q_D/\tau_{f,0}^D$ in equation (2). Hence, the contributions of states 1 and 2 to $f^D(\tau)$ and $f^E(\tau)$ differ. The most robust analysis uses both decays, mitigating difficulties in extracting lifetime distributions from multi-exponential data[18, 21].

The two-state model defines a discrete probability distribution with moments

$$\langle E \rangle = \beta E_1 + (1-\beta) E_2, \ \langle E^2 \rangle = \beta E_1^{\,2} + (1-\beta) E_2^{\,2} \quad (4)$$

The standard deviation is

$$\Delta E = \sqrt{\langle E^2 \rangle - \langle E \rangle^2} \quad (5)$$

In figure 1e, we plot ΔE versus $\langle E \rangle$ for a series of polymer simulations. This representation has not, to our knowledge, been used to present TR-FRET data, which are usually plotted as full, recovered probability distributions in *R* space. Reducing each distribution to a single point allows more data to be displayed simultaneously,

revealing patterns otherwise unseen. Using $E = \left[1 + \left(R/R_0\right)^6\right]$ rather than R confines all data points to a finite area.

The values of $\langle E \rangle$ and ΔE for different polymer regimes occupy different regions of figure 1e, giving it the "flavor" of a "phase diagram". Predictions from simulations (supplement S5) are shown for: (i) rigid rods (magenta and black), (ii) wormlike chains (red), (iii) Gaussian chains (green), and (iv) a hypothetical polymer that switches instantaneously between E=0 and E=1 states (largest possible fluctuations, blue). Colored triangles correspond to fluorescence decays of the same color in figures 1b and 1c.

For rigid rods (i) with rapidly rotating fluorophores ($\tau_r << \tau_f$), $\Delta E = 0$ (cyan triangle). For slower rotational diffusion ($\tau_r \approx \tau_f$), a small ΔE is expected (magenta). Rigid rods with tethers (ii) and $\tau_r \approx \tau_f$ are simulated by varying the distance between D and A attachment points and keeping r^{tether} constant (solid black).

The wormlike chain regime lies above the rigid rod regime (iii). An example with L=12 nm (varying l_P), $\tau_r \approx \tau_f$, and tethers is shown (solid red). Removing tethers decreases ΔE slightly (dotted red).

The Gaussian chain model (iv) provides an upper limit on the width of distance distributions from soft polymer fluctuations. With $\tau_r \approx \tau_f$ and tethers (varying Ll_P) we obtain the solid green curve. Fluorophore tethers are less important near the Gaussian chain regime, only shifting $\langle E \rangle$ and ΔE along the Gaussian chain curve. Contributions to ΔE from fluctuations in κ^2 and distance do not add linearly. For $\tau_r << \tau_f$, ΔE *increases* (dotted green); the opposite is true for the rigid rod (dotted black);

Experimental data (figure 3) lie between the rigid rod regime (solid black) and the largest possible fluctuations (solid blue). Data points above the Gaussian chain regime

(solid green) do not describe simple polymer behavior. Instead, they may indicate fast fluctuations between two states, involving crossing an energy barrier.

Nanosecond ALEX (nsALEX):

Using nsALEX, slowly inter-converting subpopulations (with respect to residence times) are "digitally sorted", and the conversion of TR-FRET measurements to *E* distributions is done for fast fluctuations within these subpopulations.

Individual molecules are sorted using the photon bursts detected as they diffuse through a focused laser excitation volume (figure 2a) while the donor excitation laser D_{exc} and the acceptor excitation laser A_{exc} are rapidly alternated. Originally, 25-100 μs alternation periods were obtained (μsALEX)[22]. In nanosecond ALEX (nsALEX) we use time division multiplexing (TDM) of interlaced, picosecond pulses from two synchronized mode-locked lasers to obtain 14.7 ns alternation periods in conjunction with fluorescence lifetime measurements (figure 2b). We use four detection channels, dividing emission by polarization and by *D* and *A* spectrum (supplement S1).

For each photon burst, we count $F^{D_{\text{em}}}$, the number of *D* photons, $F_{D_{\text{exc}}}^{A_{\text{em}}}$, the number of *A* photons excited by D_{exc}, and $F_{A_{\text{exc}}}^{A_{\text{em}}}$, the number of *A* photons excited by A_{exc}. The distance-dependent energy transfer efficiency ratio *E* is[23]

$$E = F_{D_{\text{exc}}}^{A_{\text{em}}} \Big/ \left(F_{D_{\text{exc}}}^{A_{\text{em}}} + F^{D_{\text{em}}} \right), \tag{6}$$

The *A* emission due to FRET is separated from the *A* emission due to A_{exc} laser; hence, we calculate the ALEX-specific stoichiometric ratio *S*,

$$S = \left(F^{D_{\text{em}}} + F_{D_{\text{exc}}}^{A_{\text{em}}} \right) \Big/ \left(F^{D_{\text{em}}} + F_{D_{\text{exc}}}^{A_{\text{em}}} + F_{A_{\text{exc}}}^{A_{\text{em}}} \right) \tag{7}$$

S affords molecular sorting into a *D*-only sub-population, an *A*-only sub-population, and a sub-population containing both *D* and *A* on the same chain (or two interacting chains). This sub-population can be further sorted according to *E*. Distance distributions that fluctuate slower than the photon burst durations are distinguished in two-dimensional (2D) E-S histograms as "static" species with different values of *E* and *S* (figure 2c).

nsALEX provides TR-FRET and TR-FA data unavailable using μsALEX. Analysis of multi-exponential fluorescence lifetime decays is not possible for single diffusing molecules since only up to ~100 photons are detected per photon burst. However, on a sub-population basis[24], such analysis is possible, probing distance distributions that fluctuate on time scales down to the fluorescence lifetime (~1 ns).

Single-molecule bursts are sorted into different species based on specific ranges in the 2D *E*-*S* histogram (figure 2c), and fluorescence lifetime histograms are formed using the photons from all bursts of each species (figures 2d-2g). The fluorescence lifetime decays of subpopulations with *D*-only, *A*-only, and high- and low-FRET are clearly distinguished. There are two lifetime decays in the *A* channels arising from D_{exc} and A_{exc}, giving a total of six lifetime decays over all channels. A global fit of all six lifetime decays using the two-state model, enhanced to extract rotational diffusion timescales τ_r^D and τ_r^A and to account for temporal response of the photodetectors (supplement S4), provides information about a selected sub-population. Corrections for photons from background and from other diffusing species that "leak" into these histograms are described in supplement S3. The global fit allows information from the donor and acceptor decays to constrain lifetimes, donor leakage, and direct excitation of *A* while extracting information on FRET and FA, greatly improving confidence in the fitted parameters.

dsDNA:

A series of dsDNA fragments was synthesized, each labeled with TMR as *D* on the 5' end of DNA, and internally labeled with Alexa 647 as *A* at five positions, with 8,13,18,23, and 28 base pair separations (figure 3, black squares; supplement S2). For rigid rod simulations accounting for fluorophore tethers (solid black line), the value $r^{\text{tether}} = 1.25$ nm was chosen to match ΔE for the dsDNA samples with 8, 13, and 18 bp *D-A* separations (supplement S5.4). Previous measurements of r^{tether} using FRET on dsDNA are in the range of 0.7 to 1.3 nm (values converted to r^{tether})[25, 26]. Points representing larger distances, in the range $0.0 < \langle E \rangle < 0.5$, deviate strongly from the simulations.

Figure 4 shows the distance dependence of $\langle E \rangle$ as a function of *D-A* separation (distance between the fluorophore attachment points) for the dsDNA series (black squares). As discussed by Lee et al. for dsDNA[27] and by Schuler et al. for poly-proline[28], the distance dependence of single molecule FRET measurements deviates from the predictions by Förster theory when ignoring tethers and using $\langle \kappa^2 \rangle = 2/3$ (red line, figure 4). Accounting for tether fluctuations and rotational diffusion (black line, figure 4) improves agreement between theory and experiment. The black and red lines are not fits, but calculations using independent spectroscopic data (R_0 = 6.9 nm)[29].

Deviations from the simulations in figures 3 and 4 for high *E* may be explained by the use of only one rotational diffusion time scale in the simulation (overestimating the effects of κ^2). For low *E*, intrinsic, possibly sequence-dependent bending of dsDNA on a length scale $<< l_P$ may contribute to the larger than expected fluctuations, supporting recent observations that short dsDNA fragments cyclize more quickly than expected[30]. For dsDNA, the wormlike chain model does not describe well the case $L << l_P$; there is a need for a new model to describe this range.

ssDNA:

Due to its charge density and flexibility, ssDNA is strongly affected by ionic strength. nsALEX is well-suited to the study of short polyelectrolytes in very dilute conditions, a regime inaccessible using standard methods[5]. We studied a 30-base, poly(dT) ssDNA labeled at the 3' end with TMR and at the 5' end with Alexa 647 (supplement S2) as a function of ionic strength. [NaCl] was varied over 3 orders of magnitude in a 20 mM tris buffer (red circles, figures 3). As [NaCl] increases, increasingly compact conformations and larger fluctuations are observed. The grid of wormlike chain simulations converts $\langle E \rangle$ and ΔE into L and l_p.

Below [NaCl]=1 M, the properties of poly (dT) match a wormlike chain with L=12±1 nm and l_p varying from 6 nm to 2.4 nm. The average base-to-base distance is h=0.40±0.03 nm ($L = N_{\text{monomers}} h$, N_{monomers} is the number of monomers). At 1-2 M NaCl, ΔE rises above the L=12 nm wormlike chain model, approaching the Gaussian chain regime (l_p decreases to ~.9 nm). This deviation may be related to calculations showing that the wormlike chain distribution is not always an appropriate model for polyelectrolytes[31].

For poly (dT), values used for h are generally in the 0.5-0.7 nm range, with most near 0.6 nm[32-34]. For our ssDNA, this gives a value for L between 15 and 21 nm (dashed and dash-dot red lines in figure 3). However, our result of L=12±1 nm is consistent with a model based on NMR measurements of the UpU dimer and light scattering measurements of poly(rU), where h=0.44 nm[35]. Although large-scale stacking was not observed[36], the shortened h was attributed to "residual stacking" conformers interspersed with extended, unstacked conformers. We note that our determination of h=0.40±0.03 nm is sensitive to finite chain length effects, giving more direct measurement of L and l_p.

The dependence of l_P on ionic strength is shown in figure 5. The values are larger than previous results (due to decreased *L*), which lie in the range of 0.75 nm to 4 nm[34, 37-39] (variation primarily due to differences in buffer conditions and sequence). In theoretical models, the persistence length is divided into intrinsic l_P^0 and electrostatic l_P^{el} components, $l_P = l_P^0 + l_P^{\mathrm{el}}$. In OSF (Odijk-Skolnick-Fixman) theory for charged wormlike chains[40, 41] and some extensions to flexible polymers[42-44], $l_P^{\mathrm{el}} \propto I^{-1}$, where I is the ionic strength (green fit). Others predict $l_P^{\mathrm{el}} \propto I^{-1/2}$ (red fit)[45, 46]. The dependence of l_P on ionic strength is better fit with $l_P^{\mathrm{el}} \propto I^{-1/2}$ dependence, similar to Ref. [39]. We obtain a value of l_P^0 of 0.6±0.2 nm.

Previously, discrepancies between measured $\langle E \rangle$ for poly (dT) and the wormlike chain model were explained as fast diffusion of the polymer ends during the fluorescence lifetime of the donor, increasing $\langle E \rangle$ (ΔE not measured)[33]. However, we measure full Gaussian chain distribution widths at high salt; fast end-to-end diffusion would decrease ΔE below this level. Additionally, with the slowing of intrachain diffusion found for nucleic acids[47] and polypeptides[48], simulations based on Rouse-Zimm models indicate minimal changes in $\langle E \rangle$ and ΔE (supplement S5). Dynamic averaging causing higher $\langle E \rangle$ is therefore unlikely.

Unfolded protein:

The third biopolymer examined by nsALEX was unfolded Chymotrypsin Inhibitor 2 (CI2) (two-state folder). In protein folding, distance distributions fluctuate over many time scales, including those of the overall folding and unfolding processes and of the faster fluctuations within the unfolded state[3]. We investigated the distance distributions of the unfolded state of four CI2 mutants with the following positions for *D* and *A*: 1-53 ($\Delta = 53$ a.a.), 1-60 ($\Delta = 60$ a.a.), and 1-53 with an 8 amino acid loop insertion (denoted 1-53+8, $\Delta = 61$ a.a.), all with Alexa 532 as *D* and Alexa 647 as *A*

($R_0 = 6.2$ nm; supplement S2). For the fourth, we used a previously studied protein[8], labeled at position 1 with TMR as *D* and position 40 with Cy5 as *A* ($\Delta = 40$ a.a., $R_0 = 6.9$ nm).

At high denaturant concentrations (6 M GdnCl; green triangles, figure 3), the ΔE for the CI2 mutants labeled with Alexa 532 and Alexa 647 are close to the Gaussian chain limit. $\langle E \rangle$ for the 1-53, 1-53+8, and 1-60 mutants decreases from 0.47 to 0.45 to 0.39, concurrent with increasing chain length. Increasing the chain length from $\Delta = 53$ to $\Delta = 60$ and $\Delta = 61$, Gaussian chain simulations predict a shift in $\langle E \rangle$ from 0.48 to 0.42, consistent with the data within error. For the 1-40 mutant, changing R_0 (for TMR-Cy5 *D*-*A* pair) and adjusting the simulation to match $\Delta = 40$, $\langle E \rangle$ increased to 0.63, matching the experimental value of $\langle E \rangle = 0.61 \pm 0.02$ (5M GdnCl; cyan hexagon). These results suggest that unfolded CI2 is unstructured[49] and flexible, even more than ssDNA! Since ΔE approaches the Gaussian chain limit, the fluctuation timescales of unfolded CI2 are longer than the fluorescence lifetime (consistent with previous studies indicating timescales >40 ns[9]).

In principle, the foregoing measurements are possible using bulk TR-FRET measurements; single-molecule sorting has only eliminated *D*-only and *A*-only species due to incomplete labeling, and allowed the study of polyelectrolytes at very low concentrations. Beyond these technical improvements, nsALEX also affords the detailed study of unfolded CI2 at lower denaturant concentrations, in the presence of a significant subpopulation of folded proteins.

Figures 6a and 6b show ΔE , $\langle E \rangle$ vs. [GdnCl] and ΔE vs. $\langle E \rangle$ for the unfolded subpopulation of the 1-53 mutant ([GdnCl] = 3, 3.5, 4, 5 M, blue diamonds). As seen in other single molecule studies of two-state folders[8-10], $\langle E \rangle$ increases with decreasing denaturant; this increase appears simultaneously with a significant folded

subpopulation. This was attributed to hydrophobic chain collapse. Here, we observe that ΔE increases above the Gaussian chain limit (figure 6b; 3.5 M and 3 M data points; supplement S3 describes corrections for multi-molecular events). An alternative explanation to the $\langle E \rangle$ increase could involve the changes in the conformational energy landscape that lead to protein folding. It is tempting to speculate that the increase in ΔE above the Gaussian limit indicates transient, unsuccessful attempts to cross the folding barrier, manifested by large conformational fluctuations (larger than simple chain fluctuations). Much work remains to test this idea, including the direct measurement of the fluctuations time scales[50].

Our understanding of biopolymers in general, and polyelectrolytes and protein folding in particular, stands to benefit from the ability demonstrated here to sort molecules into sub-populations and probe their distance distributions. The combination of single-molecule sorting with sub-ensemble averaging provides benefits greater than the individual advantages of each approach alone.

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com).

We thank Ashok Deniz for providing the 1-40 labeled mutant of CI2. We thank Xavier Michalet, Joseph Rudnick, and Roya Zandi for helpful discussions. This work was funded by the National Institutes of Health Grant #1R01-GM65382 (to S. W.) and was also performed (T.A.L.) under the auspices of the U. S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

Figure Captions

Figure 1: Extraction of FRET efficiency *E* and polymer distance distributions using TR-FRET. **a.** Four polymer and fluorophore rotational diffusion regimes. **Cyan:** a rigid rod with fast fluorophore rotation leading to $\langle\kappa^2\rangle = 2/3$ ($\tau_r \ll \tau_f$). **Magenta:** a rigid rod with slow rotational diffusion resulting in κ^2 fluctuations ($\langle\kappa^2\rangle < 2/3$; $\tau_r \approx \tau_f$). **Black**: a rigid rod polymer where rotating fluorophores are attached via flexible tethers. **Green:** Gaussian-chain model where rotating fluorophores are linked via flexible tethers. **b**. Donor (*D*) fluorescence lifetime curves. The parameters used are from dsDNA and ssDNA: $R_0 = 6.9$ nm, $\tau_{f,0}^D = 3.1$ ns, and $\tau_f^A = 1.6$ ns. When $\tau_r \approx \tau_f$, $\tau_r^D = 2.8$ ns, and $\tau_r^A = 2.4$ ns (each example has $\langle E \rangle$=0.5). **c**. Acceptor (*A*) fluorescence lifetime curves. **d**. The full distributions in FRET efficiency *E* are approximated with a discrete distribution with two *E* values. Black line: *E* distribution for Gaussian chain. Red lines: discrete two state *E* distribution with variable amplitudes. Inset: Same distributions in *R* space. **e**. *E* distribution simulations of *D-A* labeled polymers summarized in plots of standard deviation ΔE versus mean efficiency $\langle E \rangle$. (i) a rigid rod with $\tau_r \approx \tau_f$ and tethers (solid black), with $\tau_r \approx \tau_f$ and no tethers (dashed black), and with $\tau_r \ll \tau_f$ and tethers (dotted black); (ii) a wormlike chain with L=12 nm and varying l_P: with $\tau_r \approx \tau_f$ and tethers (solid red), with $\tau_r \approx \tau_f$ and no tethers (dotted red); (iii) a Gaussian chain with varying Ll_P: with $\tau_r \approx \tau_f$ and tethers (solid green), with $\tau_r \ll \tau_f$ and tethers (dotted green); and (iv) a hypothetical polymer with the largest possible fluctuations $\Delta E = \sqrt{\langle E \rangle (1 - \langle E \rangle)}$ (solid blue). $\langle E \rangle$ and ΔE corresponding to curved in **B** and **C** shown as colored triangles.

Figure 2: TR-FRET curves for subpopulations extracted using nsALEX (experimental setup: supplement S1). **a**. Detected photons are timed with 50 ns resolution, and classified by spectrum (*D* (yellow) or *A* (red)) and polarization. Photons delimited by dotted lines show example photon burst from single molecule diffusing through laser spot. **b.** Interlacing pulses from two lasers, D_{exc} (green; excites *D*) and A_{exc} (red; excites *A*), with a fixed delay provides a 14.7 ns alternation period for ALEX. The time delay between the detected photons and the D_{exc} laser pulse is measured with ~500 ps resolution, providing fluorescence lifetime and the ability to classify photons as due to D_{exc} or A_{exc} (whether time delay τ is before or after the A_{exc} pulse). **c.** Photon burst histogram resulting from single molecules of *D-A* labeled 1-60 CI2 mutant diffusing through laser spot in 2.4 M GdnCl. *E* and *S* are calculated for each burst and placed in histogram. Four species are detected, and their corresponding time-resolved fluorescence decay curves are extracted, shown in **c-g**: parallel *D* decay (black); perpendicular *D* decay (green); parallel *A* decay (red); perpendicular *A* decay (cyan) **d.** Proteins with *D* only emit only after D_{exc} (leakage of *D* into *A* channel removed for clarity). **e.** Proteins with *A* only emit only after A_{exc} (direct excitation of *A* by D_{exc} removed for clarity). **f.** Unfolded proteins labeled with *D* and *A* emit *D* and *A* fluorescence after D_{exc} pulse (ratio of intensities and lifetimes depend on FRET efficiency), and emit *A* fluorescence after A_{exc}. **g.** Folded proteins labeled with *D* and *A* emit similarly to case **f**, except with a higher relative intensity of *A* compared to *D* after the D_{exc} pulse, and with a shorter *D* lifetime, indicating a higher *E* due to shorter average distance.

Figure 3: ΔE vs. $\langle E \rangle$ derived from experiments on *D-A* labeled biopolymers: dsDNA with 8,13,18,23, and 25 base pair *D-A* separations (black squares);

ssDNA at varying salt concentration (red circles). 1-53, 1-53+8 and 1-60 CI2 mutants at 6M GdnCl (green triangles), 1-40 CI2 mutant at 5M GdnCl (cyan hexagon). Solid black, red, and green lines same as in **A**; red dotted line: wormlike chain with *L*=8 nm; red dashed line: wormlike chain with *L*=16 nm; red dash-dot line: wormlike chain with *L*=20 nm.

Figure 4: $\langle E \rangle$ versus *D-A* distance. Experimental results for dsDNA with 8,13,18,23, and 25 base pair *D-A* separations (black squares). Base-pair separation converted to distance according to helix rise of 0.34 nm per base pair. Simulation based on FRET model with $R_0 = 69$ Å and $\langle \kappa^2 \rangle = 2/3$ (solid red) and simulation accounting for tethers and slower rotational diffusion ($\tau_r^D = 2.8$ ns and $\tau_r^A = 2.4$ ns ; solid black) are shown.

Figure 5: Red circles: $\langle E \rangle$ for 30 base poly (dT) versus ionic strength ([NaCl]+[Tris buffer]). Solid green line: fit to model, $l_P = l_P^0 + l_P^{\mathrm{el}}$ with constant, intrinsic persistence length l_P^0 and electrostatic contribution, $l_P^{\mathrm{el}} = I^{-1}$. Solid red: fit to model with $l_P^{\mathrm{el}} = I^{-1/2}$.

Figure 6: a. ΔE versus $\langle E \rangle$ for the unfolded sub-population of 1-53 CI2 mutant at varying denaturant concentration ([GdnCl] = 3, 3.5, 4, 5 M, blue diamonds). Simulations for the Gaussian chain with dye tethers varying Ll_P are shown (solid green). **b.** Same data as in **a,** plotted as $\langle E \rangle$ (blue diamonds) and ΔE (red squares) versus [GdnCl] for experimental results with CI2 (1-53 labeling, unfolded sub-population).

References:

1. de Gennes, P. G. Brownian motions of flexible polymer chains. Nature 282, 367-370 (1979).

2. Barrat, J.-L., and Jean-Francois Joanny. in Advances in Chemical Physics (ed. Rice, I. P. a. S. A.) 1-65 (John Wiley & Sons, 1996).

3. Eaton, W. A. et al. Fast kinetics and mechanisms in protein folding. Annual Review of Biophysics and Biomolecular Structure 29, 327-359 (2000).

4. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. Nature Structural Biology 4, 10-9 (1997).

5. Cannavacciuolo, L. & Pedersen, J. Properties of polyelectrolyte chains from analysis of angular correlation functions. Journal of Chemical Physics 117, 8973-8982 (2002).

6. Stryer, L. Fluorescence energy transfer as a spectroscopic ruler. Annual Review of Biochemistry 47, 819-46 (1978).

7. Haas, E., Wilchek, M., Katchalski-Katzir, E. & Steinberg, I. Z. Distribution of end-to-end distances of oligopeptides in solution as estimated by energy transfer. Proceedings of the National Academy of Sciences of the United States of America 72, 1807-11 (1975).

8. Deniz, A. A. et al. Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. Proceedings of the National Academy of Sciences of the United States of America 97, 5179-84 (2000).

9. Schuler, B., Lipman, E. A. & Eaton, W. A. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. Nature 419, 743-7 (2002).

10. Lipman, E. A., Schuler, B., Bakajin, O. & Eaton, W. A. Single-molecule measurement of protein folding kinetics. Science 301, 1233-5 (2003).

11. Rothwell, P. J. et al. Multiparameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 reverse transcriptase:primer/template complexes. Proc Natl Acad Sci U S A 100, 1655-60 (2003).

12. Doi, M. & Edwards, S. F. The theory of polymer dynamics (Clarendon Press; Oxford University Press, Oxford Oxfordshire, New York, 1986).

13. Thirumalai, D. & Ha., B. Y. in Theoretical and mathematical models in polymer research: modern methods in polymer research and technology (ed. Grosberg, A.) 283 (Academic Press, Boston, 1998).

14. Gobush, W., Yamakawa, H., Stockmayer, W. & Magee, W. Statistical mechanics of wormlike chains. I. Asymptotic Behavior. The Journal of Chemical Physics 57, 2839-2843 (1972).

15. Wilhelm, J. & Frey, E. Radial distribution function of semiflexible polymers. Physical Review Letters 77, 2581-2584 (1996).

16. Dale, R. E., Eisinger, J. & Blumberg, W. E. The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer. Biophysical Journal 26, 161-93 (1979).

17. dos Remedios, C. G. & Moens, P. D. Fluorescence resonance energy transfer spectroscopy is a reliable "ruler" for measuring structural changes in proteins. Dispelling the problem of the unknown orientation factor. Journal of Structural Biology 115, 175-85 (1995).

18. Beechem, J. M. & Haas, E. Simultaneous determination of intramolecular distance distributions and conformational dynamics by global analysis of energy transfer measurements. Biophys J 55, 1225-36 (1989).

19. Lakshmikanth, G. S., Sridevi, K., Krishnamoorthy, G. & Udgaonkar, J. B. Structure is lost incrementally during the unfolding of barstar. Nat Struct Biol 8, 799-804 (2001).

20. Ware, W. R., Doemeny, L. J. & Nemzek, T. L. Deconvolution of fluorescence and phosphorescence decay curves. A least-squares method. The Journal of Physical Chemistry 77, 2038-2048 (1973).

21. Istratov, A. A. & Vyvenko, O. F. Exponential analysis in physical phenomena. Review of Scientific Instruments 70, 1233-57 (1999).

22. Kapanidis, A. N. et al. Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. Proc Natl Acad Sci U S A 101, 8936-41 (2004).

23. Since we use the burst analysis only to sort molecules into different species, c. f. b. c., differences in detection and quantum efficiencies of D and A, direct excitation of A by Dexc, and leakage of D into Aem,V and Aem,H are ignored in this expression. These corrections do not increase the resolving power.

24. Deniz, A. A. et al. Ratiometric single-molecule studies of freely diffusing biomolecules. Annual Review of Physical Chemistry V52, 233-253 (2001).

25. Hochstrasser, R. A., Chen, S. M. & Millar, D. P. Distance distribution in a dye-linked oligonucleotide determined by time-resolved fluorescence energy transfer. Biophys Chem 45, 133-41 (1992).

26. Parkhurst, K. M. & Parkhurst, L. J. Donor-acceptor distance distributions in a double-labeled fluorescent oligonucleotide both as a single strand and in duplexes. Biochemistry 34, 293-300 (1995).

27. Lee, N. K. et al. Towards structure from single biomolecules: accurate Förster resonance energy transfer (FRET) using fluorescence-aided molecule sorting. Biophysical Journal in press.

28. Schuler, B., Lipman, E. A., Steinbach, P. J., Kumke, M. & Eaton, W. A. private communication.

29. Clegg, R. M. Fluorescence resonance energy transfer and nucleic acids. Methods Enzymol 211, 353-88 (1992).

30. Cloutier, T. E. & Widom, J. Spontaneous sharp bending of double-stranded DNA. Mol Cell 14, 355-62 (2004).

31. Zandi, R., Rudnick, J. & Golestanian, R. Radial distribution function of rod-like polyelectrolytes. European Physical Journal E 9, 41-46 (2002).

32. Saenger, W. Principles of nucleic acid structure (Springer-Verlag, New York, 1984).

33. Murphy, M. C., Rasnik, I., Cheng, W., Lohman, T. M. & Ha, T. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. Biophys J 86, 2530-7 (2004).

34. Mills, J. B., Vacano, E. & Hagerman, P. J. Flexibility of single-stranded DNA: use of gapped duplex helices to determine the persistence lengths of poly(dT) and poly(dA). J Mol Biol 285, 245-57 (1999).

35. Hingerty, B. E., Broyde, S. B. & Olson, W. K. The poly(rU) coil: a minimum-energy model that matches experimental observations. Biopolymers 21, 1167-88 (1982).

36. Riley, M. & Maling, B. Physical and chemical characterization of two- and three-stranded adenine-thymine and adenine-uracil homopolymer complexes. J Mol Biol 20, 359-89 (1966).

37. Rivetti, C., Walker, C. & Bustamante, C. Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. Journal of Molecular Biology 280, 41-59 (1998).

38. Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. Science 271, 795-9 (1996).

39. Tinland, B., Pluen, A., Sturm, J. & Weill, G. Persistence length of single-stranded DNA. Macromolecules 30, 5763-5765 (1997).

40. Skolnick, J. & Fixman, M. Electrostatic persistence length of a wormlike polyelectrolyte. Macromolecules 10, 944-948 (1977).

41. Odijk, T. Polyelectrolytes near rod limit. Journal of Polymer Science Part B-Polymer Physics 15, 477-483 (1977).

42. Manghi, M. & Netz, R. Variational theory for a single polyelectrolyte chain revisited. European Physical Journal B 14, 67-77 (2004).

43. Everaers, R., Milchev, A. & Yamakov, V. The electrostatic persistence length of polymers beyond the OSF limit. European Physical Journal E 8, 3-14 (2002).

44. Khokhlov, A. & Khachaturian, K. On the theory of weakly charged poly-electrolytes. Polymer 23, 1742-1750 (1982).

45. Barrat, J. & Joanny, J. Persistence length of polyelectrolyte chains. Europhysics Letters 24, 333-338 (1993).

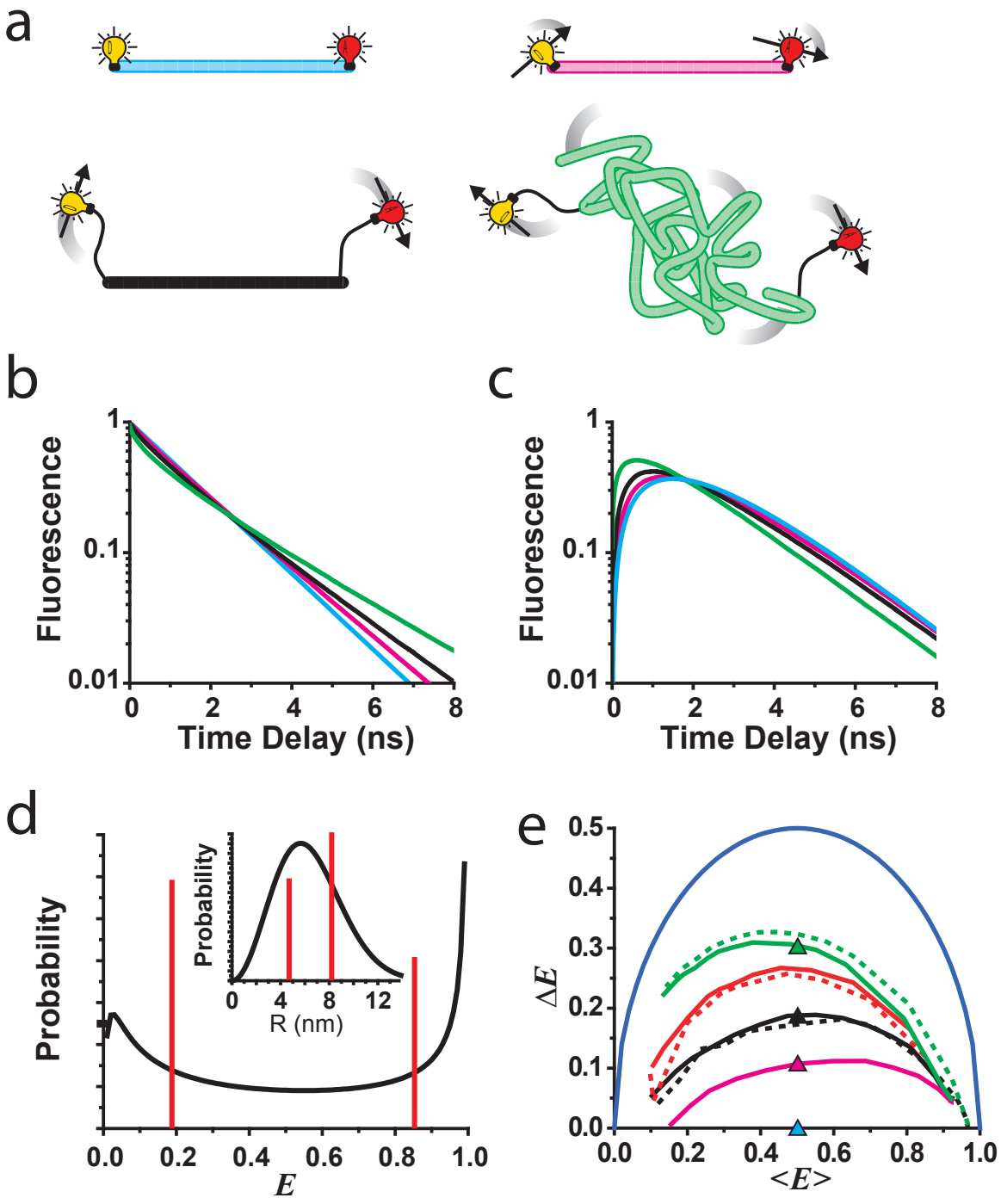
46. Ha, B. & Thirumalai, D. Electrostatic persistence length of a polyelectrolyte chain. Macromolecules 28, 577-581 (1995).

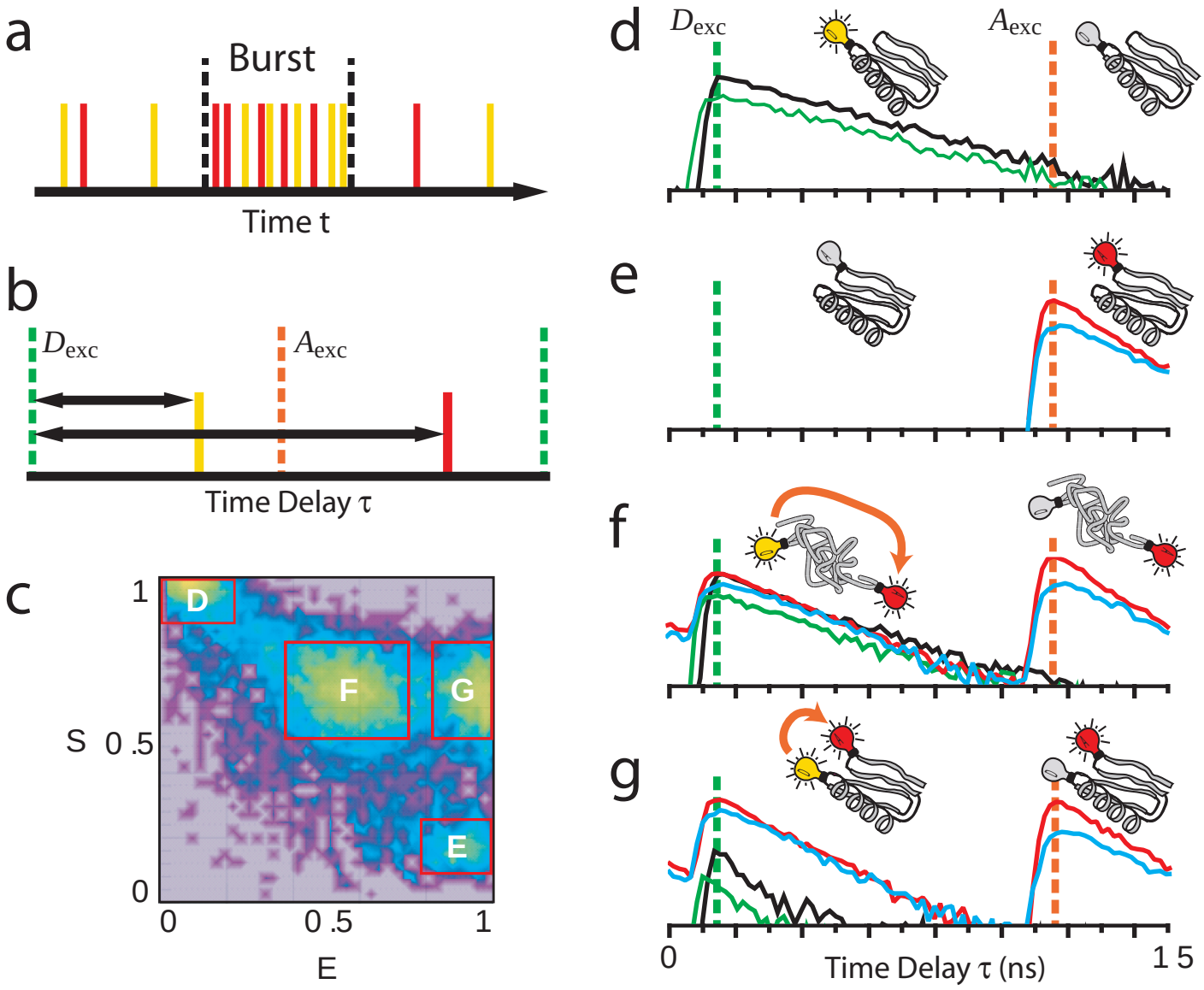
47. Wang, X. & Nau, W. M. Kinetics of end-to-end collision in short single-stranded nucleic acids. J Am Chem Soc 126, 808-13 (2004).

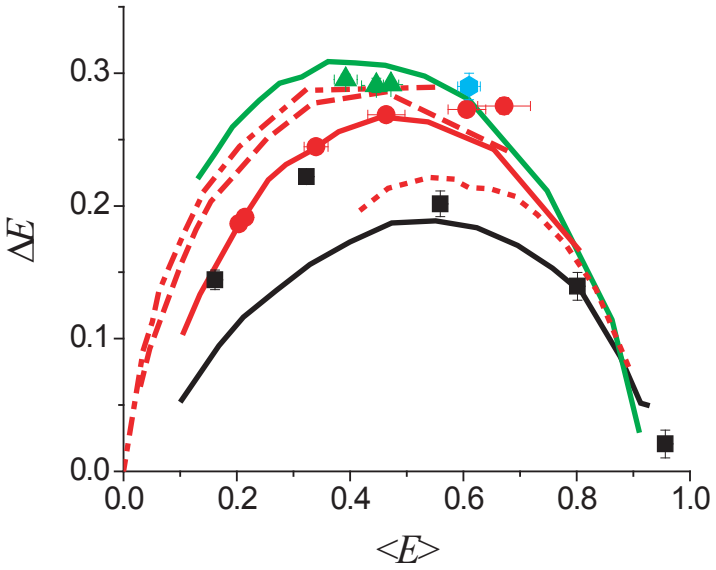
48. Lapidus, L. J., Steinbach, P. J., Eaton, W. A., Szabo, A. & Hofrichter, J. Effects of chain stiffness on the dynamics of loop formation in polypeptides. Appendix: Testing a 1-dimensional diffusion model for peptide dynamics. Journal of Physical Chemistry B 106, 11628-11640 (2002).

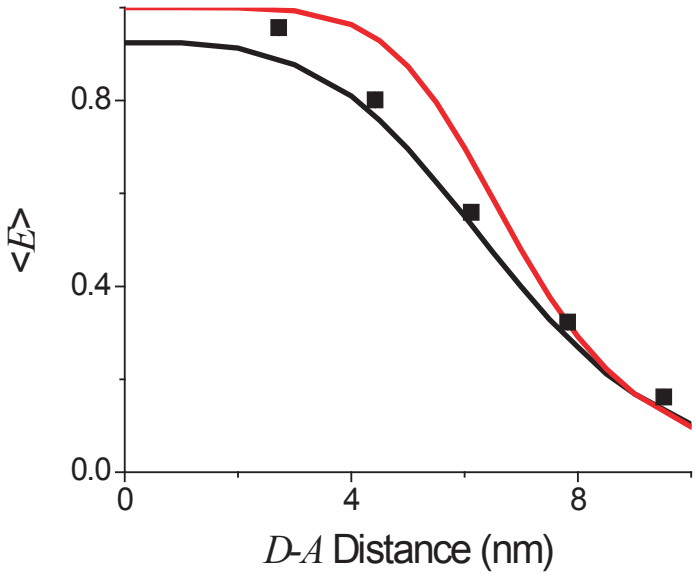
49. Kazmirski, S. L. et al. Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. Proceedings of the National Academy of Sciences of the United States of America 98, 4349-54 (2001).

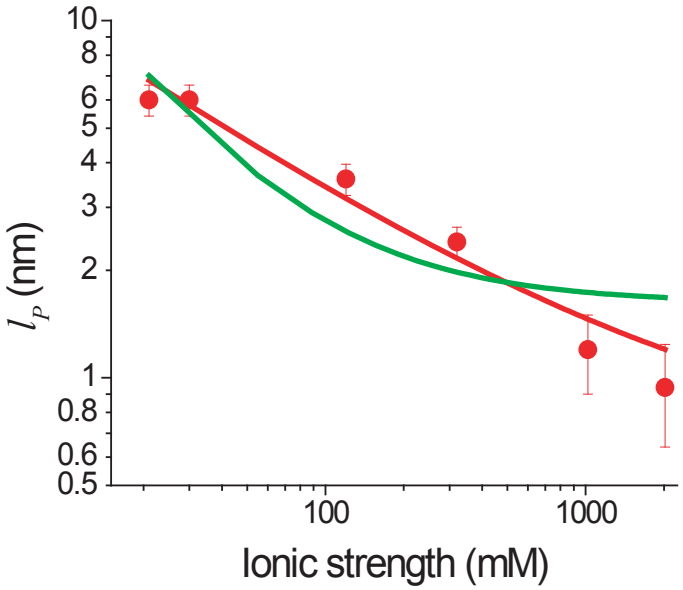
50. Berglund, A. J., Doherty, A. C. & Mabuchi, H. Photon statistics and dynamics of fluorescence resonance energy transfer. Phys Rev Lett 89, 068101 (2002).


a
b
Fluorescence
1
0.1
0.01
0
2
4
6
8
Time Delay (ns)
c
Fluorescence
1
0.1
0.01
0
2
4
6
8
Time Delay (ns)
d
Probability
0.0
0.2
0.4
0.6
0.8
1.0
E
Probability
0
4
8
12
R (nm)
e
ΔE
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.2
0.4
0.6
0.8
1.0
<E>

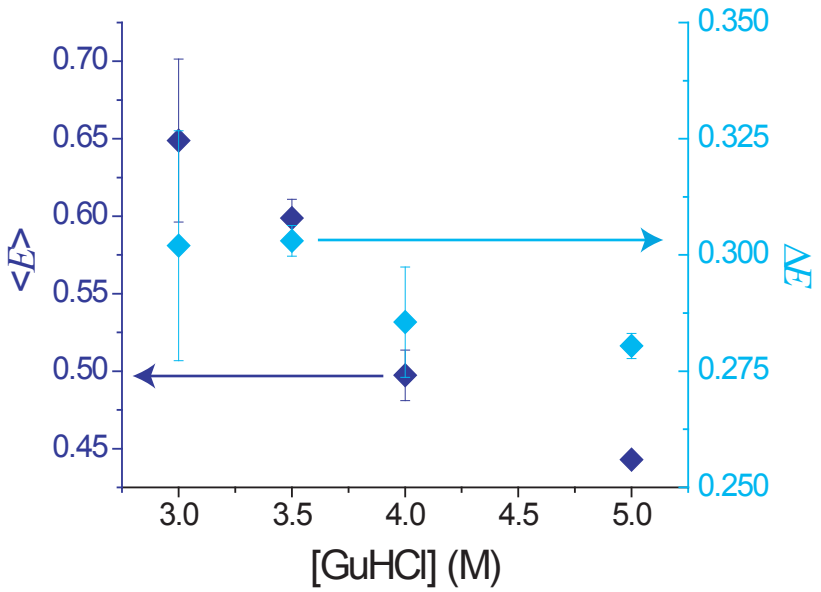

a
Burst
Time t
b
D_{exc}
A_{exc}
Time Delay τ
c
D
F
G
E
S
E
1
0.5
0
0
0.5
1
d
D_{exc}
A_{exc}
e
f
g
0
Time Delay τ (ns)
15


0.3
0.2
0.1
0.0
ΔE
0.0
0.2
0.4
0.6
0.8
1.0
<E>

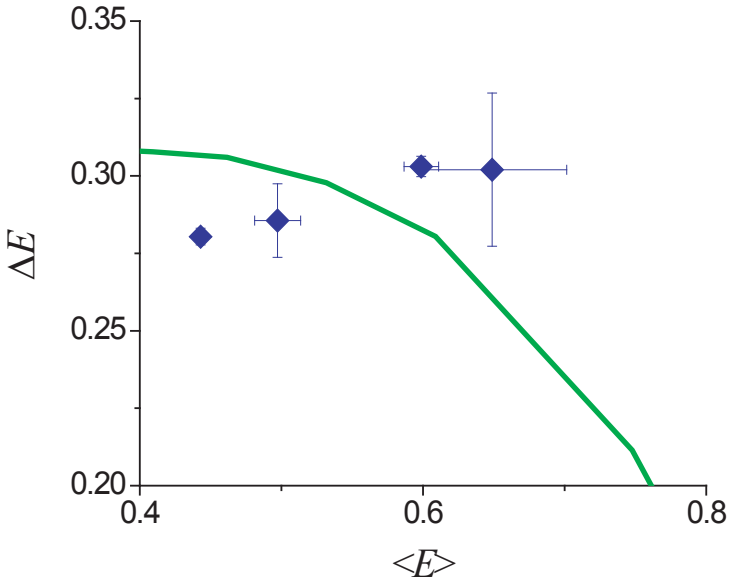

0.8
0.4
0.0
<E>
0
4
8
D-A Distance (nm)


10
8
6
5
4
3
2
1
0.8
0.6
0.5
l_P (nm)
100
1000
Ionic strength (mM)

a


0.70
0.65
0.60
0.55
0.50
0.45
<E>
0.350
0.325
0.300
0.275
0.250
ΔE
3.0
3.5
4.0
4.5
5.0
[GuHCl] (M)

b


0.35
0.30
0.25
0.20
ΔE
0.4
0.6
0.8
<E>