Sub-millisecond conformational transitions of short single-stranded DNA lattices by photon correlation single-molecule FRET

Brett Israels,[1,2] Claire S. Albrecht,[3] Anson Dang,[1] Megan Barney,[1] Peter H. von Hippel,[1] and Andrew H. Marcus[1,2,3*]

[1.] Department of Chemistry and Biochemistry, Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA

[2.] Department of Chemistry and Biochemistry, Center for Optical, Molecular and Quantum Science, University of Oregon, Eugene, OR 97403, USA

[3.] Department of Physics, Center for Optical, Molecular and Quantum Science, Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA

[*]Corresponding author: e-mail: ahmarcus@uoregon.edu

Abstract

Thermally-driven conformational fluctuations (or 'breathing') of DNA plays important roles in the function and regulation of the 'macromolecular machinery of genome expression.' Fluctuations in double-stranded (ds) DNA are involved in the transient exposure of pathways to protein binding sites within the DNA framework, leading to the binding of functional and regulatory proteins to single-stranded (ss) DNA templates. These interactions often require that the ssDNA sequences, as well as the proteins involved, assume transient conformations critical for successful binding. Here we use microsecond-resolved single-molecule Förster Resonance Energy Transfer (smFRET) experiments to investigate the backbone fluctuations of short (ss) oligo-oligo(dT)$_n$ templates within DNA constructs that can also serve as models for ss-dsDNA junctions. Such junctions, as well as the attached ssDNA sequences, are involved in the binding of ssDNA binding (ssb) proteins that control and integrate the mechanisms of DNA replication complexes. We have used these data to determine multi-order time-correlation functions (TCFs) and probability distribution functions (PDFs) that characterize the kinetic and thermodynamic behavior of the system. We find that the oligo(dT)$_n$ tails of ss-dsDNA constructs inter-convert, on sub-millisecond time-scales, between three macrostates with distinctly different end-to-end distances. These are: (*i*) a 'compact' macrostate that represents the dominant species at equilibrium; (*ii*) a

'partially extended' macrostate that exists as a minority species; and (*iii*) a 'highly extended' macrostate that is present in trace amounts. We propose a model for ssDNA secondary structure that advances our understanding of how spontaneously formed nucleic acid conformations may facilitate the activities of ssDNA associating proteins.

Significance Statement: The genetic information of living organisms is encoded as sequences of nucleic acid bases in DNA, and is protected by the thermodynamically stable secondary structure of the Watson-Crick double helix. The processing and manipulation of gene sequences by 'macromolecular machines' requires that stable segments of duplex DNA be disrupted, and that single-stranded (ss) DNA templates be transiently exposed to the binding sites of DNA associating proteins within the cellular environment. Here we elucidate some of the defining features that control the stability and dynamics of ssDNA secondary structure, using time-resolved methods to detect the presence of transient unstable conformations. Understanding the nature of these instabilities is central to elucidating the mechanisms by which ssDNA templates facilitate protein binding and function.

1. Introduction

The genome of living cells exists primarily as B-form double-stranded (ds) DNA, which is packaged in the cell nucleus as chromatin (in eukaryotes) or nucleoid bodies (in bacteria) (1). During various stages of the cell cycle, segments of single-stranded (ss) DNA are exposed by the action of the DNA replication helicase to make available the ssDNA templates required by the DNA polymerases for the synthesis of the complementary daughter strands for cell division, and to allow access of other functional and regulatory proteins to these transiently formed ssDNA sequences. Within the cell, and in functional complexes such as the replisome, these exposed ssDNA segments are coated by ssDNA binding (ssb) proteins that function to protect the genetic information from nucleases and to favorably configure the single strands for further manipulation by the 'machinery of genome expression' (2, 3). These proteins interact with ssDNA to form stable, cooperatively bound nucleoprotein filaments, in which the bound ssDNA chain segments are nearly fully extended to their maximum contour lengths permitted by the sugar-phosphate

backbones, resulting also in largely unstacking the attached bases (4, 5). Because such fully extended conformations of ssDNA are likely to be unstable in the absence of bound protein, a fundamental question arises concerning how ssb protein-ssDNA interactions evolve along a molecular reaction coordinate to form extended nucleoprotein filaments.

Unlike the cooperatively-structured and stabilized dsDNA conformation, many possible conformations can be adopted by ssDNA in solution, and the specifics of these conformations depend on a variety of interactions between the nucleic acid components with their aqueous surrounds. These interactions include base stacking and unstacking, intra-chain repulsion between adjacent backbone phosphates, counterion condensation along the polyelectrolyte ssDNA backbone, binding and orientation of polar water molecules around these counterions and the backbone phosphates, strain within the sugar-phosphate backbone and increases in configurational entropy of the backbone as a consequence of base-unstacking and the concomitant rotation around multiple single-bonds of the extended backbone (6-8).

The cumulative effect of these interactions is a free energy landscape that determines the probability with which a given ssDNA configuration will occur. In general, the coordinate space over which the free energy landscape is defined can be partitioned into distinct sub-spaces, herein referred to as 'macrostates,' which correspond to local basins of stability separated by relatively high activation barriers (9). Each macrostate consists of a 'family' of energetically-degenerate microscopic conformations (or 'microstates') that rapidly interconvert (on tens-of-nanoseconds time-scales) in a non-Markovian manner (10-12). In contrast, higher thermal barriers separate conformations associated with different macrostates, such that their interconversion occurs on tens-of-microseconds and longer time scales. The relatively high activation barriers associated with macrostate interconversion ensures that the observed dynamics of the overall system can be accurately described using a kinetic master equation in which the rate constants contain no 'hidden variables' and memory effects need not be considered (13).

In this work, we present microsecond-resolved single-molecule Förster resonance energy transfer (smFRET) experiments to monitor the conformational fluctuations of short single-stranded oligo-deoxythymidine [oligo(dT)$_n$] templates of varying length (n = 14 and 15) and polarity (3' versus 5'). We note that these constructs have been used in previous studies as model systems to monitor the dynamics of the non-cooperative and cooperative binding and assembly of

the T4 bacteriophage ssb proteins (gene product 32, or gp32) onto target ssDNA lattices (14, 15). The binding site size of the gp32 monomer is 7 nucleotides (nts) so that for the templates with a total length, n, of 14 nts there is excess room for the non-cooperative binding of a single gp32 monomer or, alternatively, precisely enough room for the cooperative assembly of a gp32 dimer. For templates with n = 15 nts, there are two possible conformations for the gp32 dimer-bound state; either at nucleotide positions 1 – 14 or 2 – 15. Furthermore, the gp32 proteins are expected to assemble onto ssDNA lattices in a polar manner, so that the assembly mechanism will depend on the polarity of the ss-ds DNA junction. Our previous studies showed that gp32 proteins bind to, and dissociate from, the ss-dsDNA junction as monomers on tens-of-milliseconds time scales, and that gp32 monomer-bound intermediates must encounter (also, on tens-of-milliseconds time scales) a 'productive' ssDNA conformation near the ss-dsDNA junction to permit a second ssb protein to form a cooperatively-bound gp32 dimer-cluster (14). By examining the sub-millisecond dynamics of the ssDNA lattice fluctuations in these constructs in the absence of proteins, we may ask whether short-lived conformational intermediates – which can serve as potential transition states for ssb binding – play a role in the assembly mechanism.

Each of the oligo(dT)$_n$ conformations that we investigated were formed within the overhanging 'tail' of a ss-dsDNA junction construct, in which the distal end of the ssDNA chain is labeled with the FRET donor chromophore Cy3 and the conjugate DNA strand is labeled, at the ss-dsDNA junction, with the FRET acceptor chromophore Cy5 (see Fig. 1). Such ss-dsDNA constructs labeled in this way were previously used by Lohman and Ha and co-workers in related studies with the E. coli SSB system using standard smFRET methods (3, 16, 17). Our experiments detect individual fluorescence photons from a continuously irradiated single molecule sample using instrumental methods (18) and data analyses developed previously (14, 19). Because these experiments resolve individual photon detection events, the dominant contribution to measurement uncertainty is the sparse statistical sampling of the ensemble-averaged signal rate (20), which varies stochastically in time due to the conformational fluctuations of the macromolecule. The time-dependent signal rates of the Cy3 donor (D) and the Cy5 acceptor (A) FRET-pair chromophores are given by $I_{D(A)}(t) = N_{D(A)}(t)/T_w$, where $N_{D(A)}$ is the number of photons detected during a fixed time window T_w (= 10 μsec) beginning at the time t. Thus, the smFRET efficiency, $E_{FRET}(t) = I_A(t)/[I_D(t) + I_A(t)]$, is a time-averaged signal, which represents an average of the macromolecular configurations detected during the sampling period T_w.

Our microsecond-resolved smFRET measurements contain both structural and kinetic information about the microscopic configurations of the oligo(dT)$_n$ templates. To extract the maximum amount of information possible from our data, we applied an 'inverse-problem' procedure for analyzing photon correlation measurements of this type (14, 19). From the photon count data stream, we determined the following three ensemble-averaged functions of the smFRET signal: (*i*) the time-averaged probability distribution function (PDF) of the mean FRET efficiency $P(E_{FRET})$; (*ii*) the two-point time correlation function (TCF), $\bar{C}^{(2)}(\tau)$, which describes the average correlations between two successive measurements separated by the time interval τ; and (*iii*) the four-point TCF, $\bar{C}^{(4)}(\tau_1, \tau_2, \tau_3)$, which describes the average correlations between four successive measurements separated by the time intervals τ_1, τ_2 and τ_3. The above functions can be used to characterize the equilibrium and kinetic properties of a multi-step conformational transition pathway at equilibrium. While the two-point TCF, $\bar{C}^{(2)}$, contains information about the characteristic relaxation times of the system, the four-point TCF, $\bar{C}^{(4)}$, contains additional information about the exchange times between pathway intermediates (19, 21-23).

We simulated the PDF and the TCFs using a master equation (24, 25) that models the system as a network of coupled macrostates, each of which is distinguished by a different mean FRET efficiency. Kinetic rate constants were varied as input parameters to the master equation to achieve the 'best fit' between simulated and experimentally-derived functions. As we discuss further below, we compared the results of a linear versus cyclical three-state network model from which we obtained optimized values for the forward and backward rate constants that connect the conformational macrostates, in addition to the probability of observing each macrostate. The optimized rate constants so obtained determine the free energy barriers that separate each macrostate (applying the Arrhenius equation), and the optimized probabilities determine the free energy minima of each macrostate (applying the Boltzmann distribution).

A major finding of this work is that all four of the oligo(dT)$_n$ templates that we investigated undergo conformational dynamics over a broad range of time scales, spanning tens-of-microseconds to hundreds-of-milliseconds. We observed that all four of these systems interconvert between three distinct macrostates: (*i*) a 'compact' and relatively stable macrostate; (*ii*) a 'partially extended' macrostate of intermediate stability; and (*iii*) a 'highly extended' macrostate of relatively low stability. The free energy minima that we determined for the three macrostates

increase rapidly with decreasing FRET efficiency, suggesting that the primary mechanism of macrostate destabilization is the loss of configurational entropy associated with increasing chain extension (10). Moreover, our results indicate the existence of transition state barriers that are intrinsic to the oligo(dT)$_n$ templates themselves, and which prolong the lifetimes of extended chain configurations relative to the more stable compact configurations. The existence of metastable extended chain conformations suggests that these short-lived species may serve as intermediates along the reaction pathway for the formation of ssb nucleoprotein clusters or filaments.

2. Materials and Methods

2.1 Single-Stranded (ss) – Double-Stranded (ds) DNA Constructs. We performed microsecond-resolved smFRET experiments on four ss-dsDNA constructs with a ssDNA oligo(dT)$_n$ overhanging 'tail' region that varied in length (n = 14 and 15) and polarity (3' and 5', see Table 1). In Fig. 1 we show a representative schematic diagram using the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct. These samples were fluorescently labeled using the Cy3/Cy5 FRET donor-acceptor chromophore pair, as indicated.

Table 1. Nucleotide base sequences for ss-dsDNA constructs studied in this work.

DNA construct	Nucleotide base sequence
3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA	5'-GTCGCCAGCCTCGCAGCCTTTTTTTTTTTTTTT/**Cy3**/-3' 3'-**biotin**/CAGCGGTCGGAGCGTCGG-**Cy5**/-5'
3'-Cy3/Cy5-oligo(dT)$_{14}$-ss-dsDNA	5'-GTCGCCAGCCTCGCAGCCTTTTTTTTTTTTTT/**Cy3**/-3' 3'-**biotin**/CAGCGGTCGGAGCGTCGG-**Cy5**/-5'
5'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA	3'-GTCGCCAGCCTCGCAGCCTTTTTTTTTTTTTTT/**Cy3**/-5' 5'-**biotin**/CAGCGGTCGGAGCGTCGG-**Cy5**/-3'
5'-Cy3/Cy5-oligo(dT)$_{14}$-ss-dsDNA	3'-GTCGCCAGCCTCGCAGCCTTTTTTTTTTTTTT/**Cy3**/-5' 5'-**biotin**/CAGCGGTCGGAGCGTCGG-**Cy5**/-3'

2.2 Sample Preparation. Sample solutions were prepared in an aqueous buffer containing 100 mM NaCl, 6 mM $MgCl_2$ and 10 mM Tris (pH 8.0), as described previously (14, 18, 26). Oxygen scavenging and triplet quenching reagents were used to extend the excited singlet state lifetime

and to minimize triplet state blinking (27, 28). Microfluidic sample chambers were constructed using fused silica quartz microscope slides, as shown in Fig. S1*A* (29). Single molecules were attached to the slide surface using a neutravidin linker, which binds strongly to biotin attached to both the ss-dsDNA constructs (Table 1) and to the poly(ethylene glycol) (PEG) layer that coats the slide surface (see Fig. 1*A*). Additional details of the sample preparation are included in the SI.

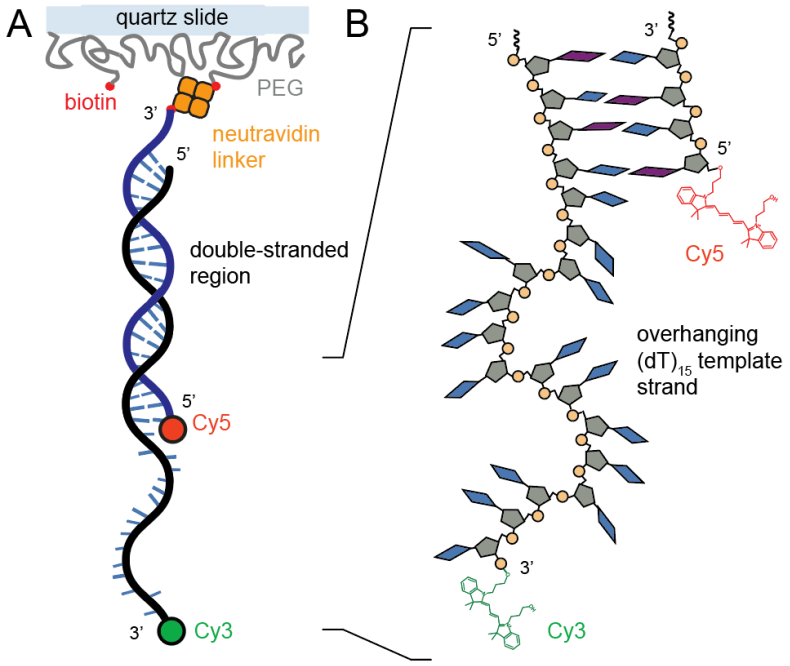

Figure 1. Schematic of one of the four single-stranded (ss) – double-stranded (ds) DNA constructs [3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA – see Table 1] used in our experiments. (***A***) The double-stranded region of the ss-dsDNA construct was attached to the surface of a fused silica quartz (SiO_2) slide coated withPEG and containing biotin-neutravidin linkers. (***B***) The Cy3 donor chromophore was attached to the distal end of the oligo(dT)$_n$ overhanging 'tail' region of the ss-dsDNA construct, and the Cy5 acceptor was attached to the opposite strand at the ss-dsDNA junction, as shown.

2.3 Microsecond-Resolved Single-Molecule (sm)FRET. We performed our smFRET experiments on the ss-dsDNA constructs using instrumentation and procedures described previously (14, 18). Additional details of the data acquisition and signal calibration are described in the SI. A schematic diagram of the instrumental setup is shown in Fig. S1*B*. The sample was illuminated using a continuous wave laser at 532 nm set up in a prism-type total internal reflection fluorescence (TIRF)

configuration. The laser beam was focused to a 50 μm diameter spot at the sample and the incident power was adjusted to 10 mW. The Cy3 donor (*D*) and Cy5 acceptor (*A*) fluorescence from a single ss-dsDNA construct was imaged through a pinhole, and spatially separated into *D* and *A* paths using a dichroic beam splitter. Individual photons from the *D* and *A* paths were detected using avalanche photodiodes (APDs), and each detection event was assigned a 'time-stamp' with 0.1 μs resolution and recorded to the computer hard disk. In post-data acquisition, we determined the time-dependent *D*(*A*) signal rates according to $I_{D(A)}(t) = N_{D(A)}(t)/T_w$, where $N_{D(A)}$ was the number of *D*(*A*) photons detected during an adjustable dwell period T_w ($\geq 10\ \mu$sec) at time t. The mean total fluorescence photon detection rate, $\bar{I}_T = \bar{I}_D + \bar{I}_A$, was typically ~15,000 – 20,000 cps. The signal 'background rate' was determined to be ~ 2,000 cps by scanning the *xy*-stage position away from the center of a molecule. Thus, the signal-to-background ratio of our experiments is ~ 10. Each single-molecule data set was recorded for a total duration of thirty seconds. From the time-dependent signal rates we determined the FRET efficiency according to $E_{FRET}(t) = I_A(t)/I_T(t)$.

3. Theoretical Background

In our experiments, we monitored the oligo(dT)$_n$ end-to-end distance R_{DA} between the *D* (Cy3) and *A* (Cy5) chromophore labels of the ss-dsDNA constructs (see Fig. 1) through measurements of the smFRET efficiency: $E_{FRET} = [1 + (R_{DA}/R_0)^6]^{-1}$. Here $R_0 = 0.211(J_{DA}\kappa^2 n^{-4}\Phi_D)^{1/6}$ is the characteristic Förster distance (in units of Å) where Φ_D is the *D* fluorescence quantum yield, n is the refractive index of the medium, J_{DA} is the overlap integral between the *D* emission and the *A* absorbance spectra, and κ^2 is the transition dipole-dipole orientation factor (30). The orientation factor $\kappa^2 = \left[\hat{\mu}_D \cdot \hat{\mu}_A - 3\left(\hat{\mu}_D \cdot \hat{R}_{DA}\right)\left(\hat{R}_{DA} \cdot \hat{\mu}_A\right)\right]^2$ accounts for the relative angles between the *D* and *A* chromophores where $\hat{\mu}_{D(A)}$ is the unit vector that points in the direction of the *D* (*A*) transition dipole moment and $\hat{R}_{DA}$ is the unit vector that points from the *D* to the *A*. For cases in which an isotropic distribution of relative *D* and *A* orientations can be assumed, which would require that at least one of the two dyes be freely rotating around its 'tether,' the orientation factor takes on the value κ^2 = 2/3 and the Förster distance $R_0 \cong 56$ Å (31, 32). In the remainder of this work, we focus on the properties of the observable E_{FRET} as an indicator of

the end-to-end distance, R_{ee} (assumed $\approx |R_{DA}|$), of the oligo(dT)$_n$ 'tail' regions of the ss-dsDNA constructs that we studied, although we return to the interpretation of the observable in terms of R_{ee} in the Conclusions.

3.1 Macrostates and Free Energy Minima. The oligo(dT)$_n$ templates can adopt a broad range of microscopic chain configurations (i.e., microstates). Assuming there are N total microstates available at a given temperature T, the probability that the ith microstate occurs is given by the Boltzmann distribution (33)

$$p_i = \frac{1}{Z} e^{-G_i/k_B T} \tag{1}$$

where k_B is Boltzmann's constant, G_i is the Gibbs free energy of the ith microstate, and $Z = \sum_{i=1}^{N} e^{-G_i/k_B T}$ is the partition function. We note that the normalized probabilities sum to unity: $\sum_{i=1}^{N} p_i = 1$. Individual microstates are expected to interconvert on tens-of-nanosecond time scales, which is much faster than the microsecond resolution of our experiments (10). Thus, the 'observed' D and A fluorescence signals each represents a time-average over the possible microstates that occur at a particular time t during the sampling period T_w (= 10 μsec). It is useful to define a 'macrostate' as a collection of pseudo-degenerate microstates that occupy a common, locally stable region of the free energy landscape. We may then regard each D and A fluorescence photon detection event as a stochastic sample of the partial-ensemble-average of microstates defined within the phase-space boundary of a single macrostate.

In our analysis, we model the probability distribution function (PDF) of E_{FRET} values as a sum of contributions from M macrostates $p(E_{FRET}) = \sum_{j=1}^{M} p_j(E_{FRET})$, where the jth macrostate has the Gaussian form $p_j(E_{FRET}) = A_j \exp\left[-\left(E_{FRET} - \langle E_{FRET}\rangle_j\right)^2/2\sigma_j^2\right]$, with mean value $\langle E_{FRET}\rangle_j$, standard deviation σ_j, and amplitude A_j. The Gaussian model for the distribution of configurational microstates (within a given macrostate) is consistent with a random-flight, finite-length polymer chain (10, 34). Macrostates that contain a relatively large number of pseudo-degenerate microstates contribute a correspondingly broad distribution of E_{FRET} values. Thus, the relative probability that the jth macrostate is present at equilibrium is given by the integrated area

of its component Gaussian $p_j^{eq} = \int_{-\infty}^{\infty} p_j(E_{FRET}) = A_j\sigma_j\sqrt{2\pi}$. We note that the sum of normalized probabilities over all macrostates is equal to unity: $\sum_{j=1}^{M} p_j^{eq} = 1$. Applying Eq. (1) to macrostates, we may thus write the free energy of the *j*th macrostate (in units of k_BT) as a function of E_{FRET}

$$\frac{G_j(E_{FRET})}{k_BT} = -\ln\left[p_j(E_{FRET})\right] = -ln\left(\frac{p_j^{eq}}{\sigma_j\sqrt{2\pi}}\right) + \frac{1}{2\sigma_j^2}\left(E_{FRET} - \langle E_{FRET}\rangle_j\right)^2 \qquad (2)$$

In the context of the above model, the free energy of the *j*th macrostate is represented as a parabolic function of E_{FRET} with vertical displacement $-ln\left(p_j^{eq}/\sigma_j\sqrt{2\pi}\right)$. The total free energy surface as a function of E_{FRET} can then be constructed by summing over the individual macrostate contributions according to $G(E_{FRET}) = -k_BT\ln\left[\sum_{j=1}^{M} p_j(E_{FRET})\right]$.

3.2 Free Energies of Activation. We assume that under equilibrium conditions, the microstates of the oligo(dT)$_n$ template are partitioned amongst M quasi-stable macrostates, and there exist conformational reaction channels that permit discreet, stochastic transitions to occur from the ith to the jth macrostate. In addition, we make the standard approximation that these transformations can be modeled as continuous processes (24, 35). We thus define the time-dependent population of the ith macrostate, $p_i(t)$, as a continuous, single-valued function that obeys the coupled, first-order ordinary differential equations $\dot{p}_i = \sum_{j\neq i=1}^{M} -k_{ij}p_i + k_{ji}p_j$. Here the forward rate constant k_{ij} is associated with the transition from macrostate i to macrostate j, and the backward rate constant k_{ji} is associated with the transition from macrostate j to macrostate i. The rate constant k_{ij} is related to the free energy of activation $G_{ij}^{\ddagger}$, which is given by the Arrhenius equation (33)

$$k_{ij} = k_{max}e^{-G_{ij}^{\ddagger}/k_BT} \qquad (3)$$

In Eq. (3), the maximum rate constant k_{max} corresponds to a barrierless process, to which all the thermally activated processes of the system are referenced.

3.3 Two-Point and Four-Point Time-Correlation Functions (TCFs). We obtain kinetic information from our time-resolved smFRET data through our analyses of two-point and four-point time-correlation functions (TCFs) (19). We define the time-dependent fluctuation of the E_{FRET} observable $\delta E_{FRET}(t) = E_{FRET}(t) - \langle E_{FRET} \rangle$, where $\langle E_{FRET} \rangle$ is the mean value determined from a time series of measurements performed on a single molecule. The two-point TCF is the average product of two successive measurements separated by the interval τ

$$\bar{C}^{(2)}(\tau) = \langle \delta E_{FRET}(0)\, \delta E_{FRET}(\tau) \rangle \tag{4}$$

In Eq. (4), the angle brackets indicate a running average over all possible initial measurement times. As we discuss further below, the function $\bar{C}^{(2)}(\tau)$ contains information about the minimum number of possible macrostates of the system, each macrostate's mean FRET efficiency value, in addition to the characteristic time scales of state-to-state interconversion. The two-point TCF may be expressed using a standard statistical mechanical model according to:

$$\bar{C}^{(2)}(\tau) = \sum_{i,j=1}^{M} \delta E_{FRET_j}\, p_{ij}(\tau) \delta E_{FRET_i}\, p_i^{eq} \tag{5}$$

Here p_i^{eq} is the probability that the ith macrostate is present at equilibrium, $\delta E_{FRET_i} = E_{FRET_i} - \sum_{i=1}^{M} p_i^{eq} E_{FRET_i}$ is the value of the fluctuation observable corresponding to the ith macrostate, and $p_{ij}(\tau)$ is the conditional probability that the system undergoes a transformation from macrostate i to macrostate j during the interval τ. Equation (5) states that $\bar{C}^{(2)}(\tau)$ is the statistically weighted average product of consecutive (two-point) observations, which are expected to occur within the interval τ as the system undergoes spontaneous transitions between the M macrostates.

We define the four-point TCF as the time-averaged product of four successive measurements separated by the intervals τ_1, τ_2, and τ_3, respectively.

$$\bar{C}^{(4)}(\tau_1, \tau_2, \tau_3) = \langle \delta E_{FRET}(0)\, \delta E_{FRET}(\tau_1) \delta E_{FRET}(\tau_1 + \tau_2) \delta E_{FRET}(\tau_1 + \tau_2 + \tau_3) \rangle \tag{6}$$

The function $\bar{C}^{(4)}(\tau_1, \tau_2, \tau_3)$ contains information about the roles of intermediates within the conformational kinetic pathways that affect the rates of interconversion between macrostates. The first term on the right-hand-side of Eq. (6) may be modeled according to:

$$\begin{aligned} &\langle \delta E_{FRET}(0)\, \delta E_{FRET}(\tau_1) \delta E_{FRET}(\tau_1 + \tau_2) \delta E_{FRET}(\tau_1 + \tau_2 + \tau_3) \rangle \\ &= \sum_{i,j,k,l=1}^{M} \delta E_{FRET_l} p_{kl}(\tau_3) \delta E_{FRET_k}\, p_{jk}(\tau_2) \delta E_{FRET_j} p_{ij}(\tau_1) \delta E_{FRET_i}\, p_i^{eq} \end{aligned} \tag{7}$$

Equation (7) expresses $\bar{C}^{(4)}(\tau_1, \tau_2, \tau_3)$ in terms of the weighted average of four-point products of the observable δE_{FRET}, in which a transition occurs from macrostate i to macrostate j during the interval τ_1, a subsequent transition occurs from macrostate j to macrostate k during the period τ_2, and a final transition occurs from macrostate k to macrostate l during the interval τ_3. We may thus view the four-point TCF as the τ_1, τ_2, τ_3-dependent weighted average of four-point transitions (i.e., elementary kinetic pathways) that connect the M macrostates. As we discuss in the next section, the two-point and four-point TCFs described by Eqs. (5) and (7), respectively, in addition to the equilibrium PDF can be modeled theoretically using a kinetic master equation.

3.4 Master Equation Analysis. In order to model the ensemble-averaged quantities discussed in the previous sections, it is necessary to determine for a given conformational reaction scheme the equilibrium distribution of macrostates, p_i^{eq}, and the time-dependent conditional probability, $p_{ij}(\tau)$, that the system will undergo a transformation from macrostate i to macrostate j during the interval τ. These probabilities are solutions to the master equation, which is the set of coupled first-order ordinary differential equations that describe the time-dependent macrostate populations.

$$\dot{\boldsymbol{p}}(t) = \boldsymbol{K}\boldsymbol{p}(t) \tag{8}$$

In Eq. (8), $\boldsymbol{p}(t) = [p_1(t), p_2(t), .., p_M(t)]$ is the time-dependent macrostate population vector, and $\dot{\boldsymbol{p}}(t)$ is its time-derivative. The $M \times M$ rate matrix $\boldsymbol{K}$ contains the rate constants k_{ij} according to

$$\boldsymbol{K} = \begin{bmatrix} -\sum_{i\neq j=1}^{M} k_{1j} & k_{21} & \cdots & k_{M1} \\ k_{12} & -\sum_{i\neq j=1}^{M} k_{2j} & & k_{M2} \\ \vdots & & \ddots & \vdots \\ k_{1M} & k_{2M} & \cdots & -\sum_{i\neq j=1}^{M} k_{Mj} \end{bmatrix} \tag{9}$$

The rate matrix $\boldsymbol{K}$ is written in the 'site-basis' of elementary kinetic steps that comprise the coupled reaction scheme. The off-diagonal elements $K_{ij} = k_{ij}$ correspond to the gain in population of macrostate j due to the transition $i \rightarrow j$. For cases in which there is no transition $i \rightarrow j$, $K_{ij} = 0$. The diagonal elements are written $K_{ii} = -\sum_{j\neq i=1}^{M} k_{ij}$, and represent the sum of all reactions that deplete population from macrostate i. Equation (8) must satisfy completeness, $\sum_{i=1}^{M} p_i(t) = 1$, and the detailed balance conditions $p_i^{eq} k_{ij} = p_j^{eq} k_{ji}$ with $p_i^{eq} = \lim_{t\rightarrow\infty} p_i(t)$. Moreover, the vector of equilibrium macrostate populations $\boldsymbol{p}^{eq}$ can be obtained by solving the master equation subject to the time-stationary boundary condition $\dot{\boldsymbol{p}}(t) = \boldsymbol{K}\boldsymbol{p}^{eq} = 0$.

Given a particular set of rate constants, we solve Eq. (8) by finding the eigenvalues, $\lambda_1, \lambda_2, \ldots \lambda_M$, and the eigenvectors, $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_M$, which satisfy the eigenvector equation $\boldsymbol{K}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$. The eigenvalues are equal to the characteristic relaxation rates of the coupled dynamical system, and the eigenvectors represent the collective modes (i.e., linear combinations of elementary transitions) associated with the characteristic relaxations. An important symmetry property of the rate matrix $\boldsymbol{K}$ is that the elements within each of its columns sum to zero (i.e., $\sum_{j=1}^{M} K_{ij} = 0$), which results in $\lambda_1 = 0$ and the remaining eigenvalues $\lambda_2, \ldots \lambda_M < 0$ (36).

A general solution to the master equation [Eq. (8)] is

$$\boldsymbol{p}_n(t) = c_1^n \boldsymbol{v}_1 e^{\lambda_1 t} + c_2^n \boldsymbol{v}_2 e^{\lambda_2 t} + \cdots + c_M^n \boldsymbol{v}_M e^{\lambda_M t} \tag{10}$$

where the constants $c_1^n, c_2^n, \dots, c_M^n$ are expansion coefficients that depend on the specific initial condition, which we indicate by the index n (25). Equation (10) can be further simplified by making the substitution $\lambda_1 = 0$ and taking the infinite-time limit, so that $\boldsymbol{p}^{eq} = c_1^n \boldsymbol{v}_1$ for all possible initial conditions.

$$\boldsymbol{p}_n(t) = \boldsymbol{p}^{eq} + c_2^n \boldsymbol{v}_2 e^{\lambda_2 t} + \cdots + c_M^n \boldsymbol{v}_M e^{\lambda_M t} \tag{11}$$

The conditional probabilities needed to evaluate the two-point and four-point TCFs [Eqs. (5) and (7), respectively] are obtained by constraining Eq. (11) using appropriate boundary conditions. The conditional probability $p_{ij}(t)$ is the probability that at time t, the population vector has unit occupancy of the jth macrostate, given that at time zero there was unit occupancy of the ith macrostate. Thus, for example, by applying the initial condition $n = i$ to Eq. (11) with $p_i(0) = 1$ and $p_{j \neq i}(0) = 0$, we may determine the expansion coefficients $c_2^i, c_3^i, \dots, c_M^i$. The conditional probability $p_{ij}(t)$ is then equal to the jth element of the population vector given by Eq. (11), according to

$$p_{ij}(t) = p_j^{eq} + c_2^i v_2^j e^{\lambda_2 t} + \cdots + c_M^i v_M^j e^{\lambda_M t} \tag{12}$$

In principle, the above procedure can be used to determine the $p_{ij}(t)$s analytically for all of M^2 possible combinations of $i, j \in \{1,2, \dots, M\}$, provided the number of states M is relatively small (19). In practice, the $p_{ij}(t)$s are determined numerically by writing Eq. (11) in the eigenbasis of the $\boldsymbol{K}$ matrix and carrying out the similarity transformation $\boldsymbol{p}_n(t) = \boldsymbol{U}\left[\boldsymbol{e}^{\boldsymbol{\lambda} t}\right]\boldsymbol{U}^{-1}\boldsymbol{p}_n(0)$, where the unitary matrix $\boldsymbol{U} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_M]$ and the matrix $\left[\boldsymbol{e}^{\boldsymbol{\lambda} t}\right]$ is diagonal with non-zero elements $\left[e^{\lambda t}\right]_{ii} = e^{\lambda_i t}$ (24). Additional details of this procedure are given in the SI.

We note that in the infinite-time limit Eq. (12) recovers the expected equilibrium population values, $p_{ij}(t \to \infty) = p_j^{eq}$. Furthermore, the conditional probabilities $p_{ij}(t)$ are each a sum of $M - 1$ decaying exponential functions of time, with decay constants equal to the eigenvalues of the $\boldsymbol{K}$ matrix. Substitution of Eq. (12) into Eq. (5) shows that the two-point TCF is also a sum

of $M-1$ exponential terms: $\bar{C}^{(2)}(\tau) = \mathcal{A}_2 e^{\lambda_2 \tau} + \mathcal{A}_3 e^{\lambda_3 \tau} + \cdots + \mathcal{A}_M e^{\lambda_M \tau}$, where the $\mathcal{A}_i$s are the relative weights of the collective relaxation processes. Similarly, the four-point TCF can be shown to be a sum of $(M-1)^2$ terms: $\bar{C}^{(4)}(\tau_1, \tau_3)_{\tau_2 \text{ fixed}} = \mathcal{A}_{22}(\tau_2) e^{\lambda_2(\tau_1+\tau_3)} + \mathcal{A}_{23}(\tau_2) e^{\lambda_2 \tau_1 + \lambda_3 \tau_3} + \mathcal{A}_{32}(\tau_2) e^{\lambda_3 \tau_1 + \lambda_2 \tau_3} + \cdots + \mathcal{A}_{MM}(\tau_2) e^{\lambda_M(\tau_1+\tau_3)}$ (19). Thus, the number of exponential terms that contribute to the two-point and four-point TCFs indicate the minimum number of macrostates M that underlie the observed dynamics.

4. Results and Discussion

For each of the four ss-dsDNA constructs that we studied (see Table 1), we recorded time series of single-molecule photon data streams, each for a duration of thirty seconds (see Sect. 2.3). The results of our analyses indicate that all four of the ss-dsDNA constructs exhibit very similar behaviors, as one might expect when focusing on conformational changes within ssDNA sequences only, although the four constructs behave differently at slower time scales in the presence of ssb (gp32) proteins (14, 15, 37). In the following discussion, which deals only with ssDNA conformational changes, we focus on the results of one of the ss-dsDNA constructs [3'-oligo$(dT)_{15}$]. We summarize our results for all four of the ss-dsDNA constructs that we studied in the SI. These additional data sets will serve as reference states in our subsequent considerations of ssDNA-ssb interactions (37).

We begin by making some qualitative observations. In Fig. 2, we show example trajectories of the donor D (Cy3, green) and acceptor A (Cy5, red) fluorescence intensities constructed from our raw single-molecule photon correlation measurements of the 3'-Cy3/Cy5-oligo$(dT)_{15}$-ss-dsDNA construct. The fluorescence intensities are calculated according to $I_{D(A)}(t) = N_{D(A)}(t)/T_w$, where $N_{D(A)}(t)$ is the number of D (A) photons recorded within a finite sampling window T_w beginning at time t. An additional example trajectory is shown in Fig. S2 of the SI. We see that the D and A fluorescence intensities fluctuate in an anti-correlated manner due to the varying inter-chromophore FRET coupling, and indicates that these signals depend on conformational changes of the 3'-oligo$(dT)_{15}$ tail region of the ss-dsDNA construct.

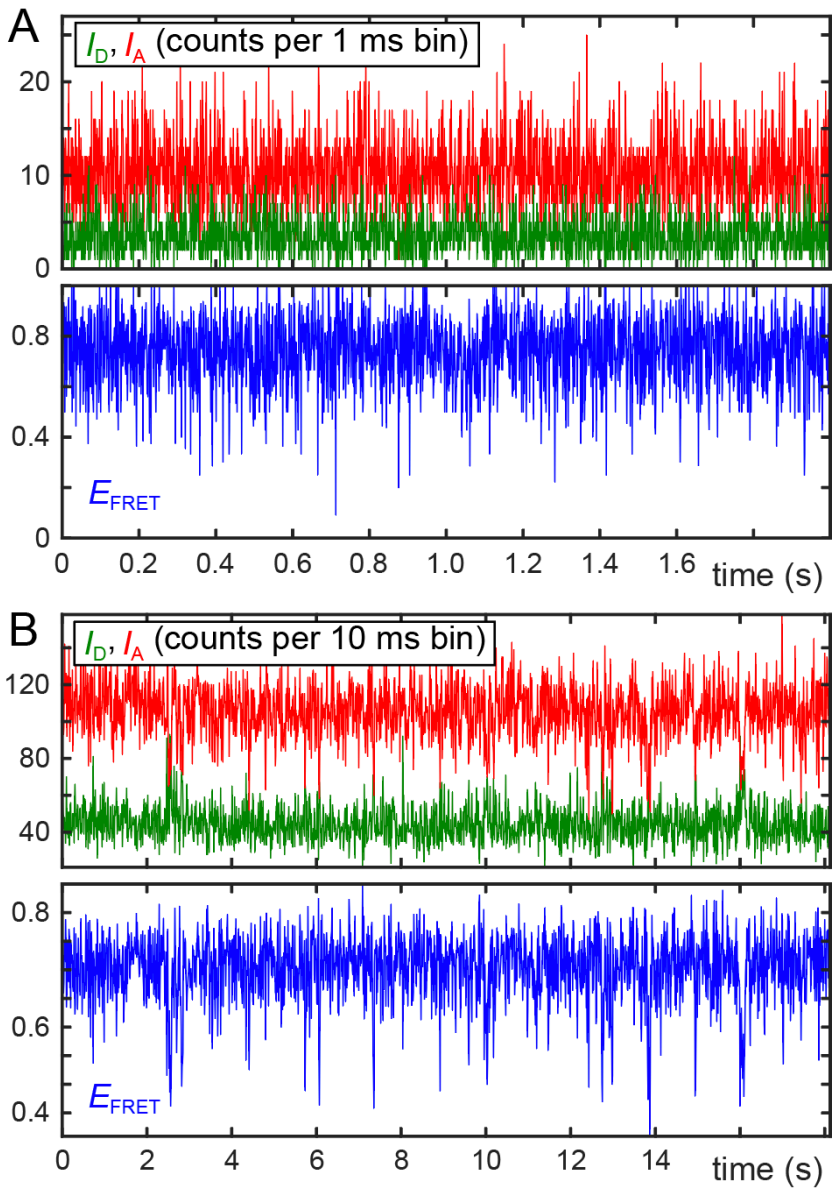

Figure 2. Example smFRET trajectories for the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct. The donor *D* (Cy3, green) and acceptor *A* (Cy5, red) fluorescence intensities, $I_{D(A)}(t) = N_{D(A)}(t)/T_w$, are shown in **panel *A*** with sampling window resolution T_w = 1 ms over a duration of ~ 2 seconds. The FRET efficiency $E_{FRET}(t)$ is plotted in the bottom panel (blue). In **panel *B*** are shown the same data using bin resolution T_w = 10 ms over a duration of ~ 18 seconds. The average Signal-to-Noise Ratio (SNR) per sampling window is $\bar{N}_T^{1/2} \cong 4$ for T_w = 1 ms, and $\bar{N}_T^{1/2} \cong 12$ for T_w = 10 ms.

The *D* and *A* intensities are used to calculated the FRET efficiency, according to $E_{FRET}(t) = I_A(t)/[I_D(t) + I_A(t)]$ (blue). We show these trajectories for two different values of

the sampling window: T_w = 1 ms (Fig. 2*A*) and T_w = 10 ms (Fig. 2*B*). For this particular photon data stream, the mean total signal flux was $\bar{I}_T = \bar{I}_D + \bar{I}_A \approx$ 15,000 s^{-1}. We note that the trajectory appears 'noisier' for the case of the T_w = 1 ms sampling window in comparison to the T_w = 10 ms window (see also Fig. S2). This is due to the measurement uncertainty associated with the statistical sampling of photons during the finite sampling window T_w at a constant low-level flux. An average Signal-to-Noise Ratio (SNR) can be assigned to the integrated signals shown in Fig. 2, SNR = $\bar{N}_T^{1/2}$, where $\bar{N}_T = \bar{I}_T T_w$ is the mean number of detected photons per sampling window. Thus, for T_w = 1 ms the average SNR $\approx \sqrt{15} \cong 4$, and for T_w = 10 ms the average SNR = $\bar{N}_T^{1/2} \approx \sqrt{150} \cong 12$. Upon further inspection of the signal trajectories shown in Fig. 2*B*, we see that the majority of FRET efficiency values are clustered near $E_{FRET} \sim 0.7$. Moreover, we observe that infrequent transitions occur to short-lived states with FRET efficiency values centered near $E_{FRET} \sim 0.5$, and even less frequent transitions occur to short-lived states with $E_{FRET} < 0.5$. In the analysis that follows, we quantify this behavior using statistical functions.

We analyzed our raw photon correlation measurements by constructing probability distribution functions (PDFs), in addition to two-point and four-point time-correlation functions (TCFs) (see Fig. 3). In constructing these functions, we compared the results from individual single-molecule data sets to check for consistency before averaging these together. Typically, ~35 individual single-molecule data sets were used to compute each function. In addition, we performed numerical simulations of the PDFs and TCFs using the master equation approach described in Sect. 3.4. Additional details of the numerical methods we employed are described in the SI.

In Fig. 3*A*, we show a normalized histogram of E_{FRET} values, which was compiled from the data streams of 35 individual single-molecule measurements of the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct. The values of the FRET efficiency were calculated by binning the *D* and *A* intensities using a sampling window of T_w = 1 ms. Each single-molecule photon data stream had an approximate mean flux $\bar{I}_T$ ~20,000 s^{-1}, and was collected over an acquisition period of 30 s. Thus, the total number of single-photon measurements used to compile the histogram is approximately ~21 million. The shape of the PDF, which is mostly peaked around $E_{FRET} \sim 0.7$, but also exhibits a shoulder at $E_{FRET} \sim 0.5$ and a tail for $E_{FRET} < 0.5$, is consistent with our qualitative observations of the time-dependent behavior of the FRET efficiency trajectory shown

in Fig. 2. The PDF cannot be modeled using a single Gaussian feature, but rather requires multiple underlying features to recover its shape. A spectral decomposition analysis suggests that the PDF is best modeled using a sum of three Gaussians, as indicated. All four of the ss-dsDNA constructs that we investigated exhibited a very similar shape to that shown in Fig. 3*A* (see Fig. S3 of the SI). We next discuss our kinetic analysis of these single-molecule data, which supports a model in which the system exists as a distribution of three quasi-stable macrostates in equilibrium. The results of this analysis provide optimized parameters for the Gaussian decomposition of the PDFs of all four of the ss-dsDNA constructs (Table S1 of the SI), and we list the results for the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct in Table 2.

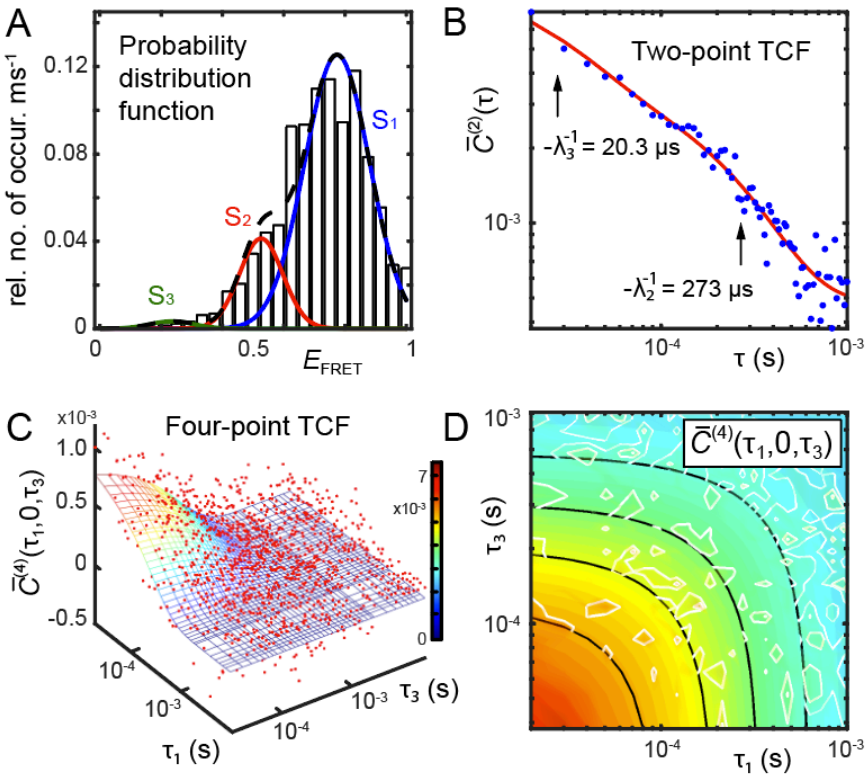

Figure 3. Experimentally derived statistical functions constructed from single-molecule photon correlation measurements of the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct. Fits to model functions based on the kinetic master equation are also shown. (***A***) The normalized histogram of E_{FRET} values using a sampling window T_w = 1 ms and 21 million photons, which is represented as a sum of three Gaussian macrostates, labeled $S_1 - S_3$. (***B***) The two-point time correlation function (TCF) of E_{FRET} values (blue points) is modeled theoretically as a weighted sum of two exponentially decaying terms (red curve) using a three-state master equation. (***C***) The four-point TCF is shown as a three-dimensional rendering (red points), which is overlayed with the theoretical model predicted by the three-state master equation

(mesh surface). (***D***) The four-point TCF in comparison to the theoretical model are shown as two-dimensional contour plots. The experimental data are shown as white contours and the model as a solid color-coded surface. The optimized values obtained from the master equation for the equilibrium PDF and the rate constant parameters are given in Table 2 and Table 3, respectively.

Table 2. Optimized parameters for the Gaussian decomposition of the FRET probability distribution function (PDF) shown in Fig. 3*A*. Parameters defined in Eq. (2). Error bars for the mean FRET efficiencies are ~0.005 (see Fig. S11 and Table S10 of the SI).

Construct	A_1	$\langle E_{FRET}\rangle_1$	σ_1	P_1^{eq}	A_2	$\langle E_{FRET}\rangle_2$	σ_2	P_2^{eq}	A_3	$\langle E_{FRET}\rangle_3$	σ_3	P_3^{eq}
3'oligo(dT)$_{15}$	3.04	0.77	0.11	0.81	1.00	0.52	0.07	0.18	0.08	0.23	0.07	0.01

We calculated experimental two-point TCFs [defined by Eq. (4)] from the time-dependent photon data streams of our single-molecule measurements (see SI for further details). For each of the ss-dsDNA constructs that we studied, we found that the two-point TCFs exhibited very similar decay dynamics, spanning the range of time scales: 10^{-6} – 10^{-3} s. These TCFs exhibited a relaxation component on the order of ~3 μs, as shown in Fig. S4 of the SI. We fit the first 20 microseconds of the TCFs to a model mono-exponential decay, and the optimized time constants for each of the ss-dsDNA constructs are listed in Table S2 of the SI. Relaxations on the time scale of ~3 μs have been observed previously in photophysical studies of the Cy3 chromophore, and are known to be associated with excited state intersystem crossing and forward photoisomerization, which undergo the reverse step of ground state recovery within ~10 μs (38-42). We therefore assign the ~ 3 μs decay in our current studies to photoisomerization of the Cy3 chromophore, and we assume that these relatively fast processes contribute to the 'time-averaged' background of the signal fluctuations that we observe on tens-of-microseconds and longer time scales, and which we interpret to be due to structural changes of the oligo(dT)$_n$ templates.

In Fig. 3*B*, we show the experimental two-point TCF over the range of time scales from 2 $\times 10^{-5}$ – 10^{-3} s, which was calculated from the same 35 single-molecule measurements of the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct used to compile the histogram shown in Fig. 3*A*. The decay is compared to a model bi-exponential fit, $\bar{C}^{(2)}(\tau) = \mathcal{A}_2 e^{\lambda_2 \tau} + \mathcal{A}_3 e^{\lambda_3 \tau} + B$, with relaxation time constants [eigenvalues of the rate matrix, Eq. (9)] λ_3^{-1}= 20.3 μs and λ_2^{-1} = 273 μs, as shown. Upon modeling the experimental decay using a tri-exponential function, we obtained

very similar values of the relaxation time constants, although the quality of the fit (as determined by the goodness-of-fit parameter R^2) did not improve (see Fig. S5 and Table S3 of the SI). A comparison between the two-point TCFs and their model fits for all four ss-dsDNA constructs are shown in Fig. S6 of the SI, and the corresponding model fit parameters are listed in Table S4 of the SI.

We calculated the four-point TCFs from our experimental data according to Eq. (6). In Fig. 3*C*, we show a three-dimensional rendering of this function, $\bar{C}^{(4)}(\tau_1, \tau_2 = 0, \tau_3)$, where we have set the intermediate time interval $\tau_2 = 0$. These data are shown overlayed with a two-dimensional model function, $\bar{C}^{(4)}(\tau_1, \tau_2 = 0, \tau_3) = \mathcal{A}_{22} e^{\lambda_2(\tau_1+\tau_3)} + \mathcal{A}_{23} e^{\lambda_2\tau_1+\lambda_3\tau_3} + \mathcal{A}_{32} e^{\lambda_3\tau_1+\lambda_2\tau_3} + \mathcal{A}_{33} e^{\lambda_3(\tau_1+\tau_3)} + D$, which is based on an optimized solution to the kinetic master equation. In Fig. 3*D*, we show the experimental and simulated four-point TCFs as a two-dimensional contour plot. We note that the three-dimensional surface of $\bar{C}^{(4)}$ exhibits pronounced convex contour lines (corresponding to negative values of $\mathcal{A}_{23} = \mathcal{A}_{32}$, see Table S5), which indicate the presence of kinetic 'bottleneck' intermediates in the reaction pathway that inhibit the rapid exchange of population among the various macrostates of the system (19). We show the four-point TCFs and their model fits to all four ss-dsDNA constructs in Fig. S7 of the SI, and the corresponding model fit parameters are listed in Table S5 of the SI. The combined statistical functions shown in Fig. 3 serve to constrain a multi-parameter optimization of solutions to the master equation.

As discussed in Sect. 3.4, the presence of two decay components in the two-point TCFs implies that there must be at least three conformational macrostates between which the oligo(dT)$_n$ templates interconvert. Moreover, a three-state model is consistent with the histogram of E_{FRET} values (see Fig. 3*A*), in which the underlying Gaussian features, labeled $S_1 - S_3$, may be assigned to a 'compact' macrostate, a 'partially- extended' macrostate, and a 'highly-extended' macrostate, respectively. The compact macrostate S_1 has the largest mean E_{FRET} value, since it contains chain configurations with relatively short $D - A$ chromophore distances. The partially-extended macrostate S_2 and the highly-extended macrostate S_3, on the other hand, have smaller mean E_{FRET} values, indicating chain configurations with correspondingly larger $D - A$ separations.

We therefore adopted a kinetic network scheme with $M = 3$ macrostates, as depicted in Fig. 4*A*, to model our results using the master equation [Eq. (8)]. The solutions to the three-state master equation can be written analytically in terms of the rate constants of elementary reactive steps (19).

We performed numerical simulations of Eq. (8) by searching the parameter space of input rate constants and FRET efficiency values, as described in our prior work (14, 19). The solutions to the master equation are the time-dependent conditional probabilities $p_{ij}(t)$ [Eq. (12)] and the equilibrium distribution of states $\boldsymbol{p}^{eq} = \left(p_1^{eq}, p_2^{eq}, p_3^{eq}\right)$ [Eq. (11)], which may be used to simulate expressions for the PDF and the two-point and four-point TCFs described in Sect. 3.3. We performed an iterative search of the parameter space in which the statistical functions $p(E_{FRET})$, $\bar{C}^{(2)}(\tau)$ and $\bar{C}^{(4)}(\tau_1, 0, \tau_3)$, shown in Fig. 3 for the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct, were calculated for a given set of input rate constants and FRET efficiency values, and compared to the experimentally-derived functions. The agreement between simulated and experimental functions was quantified using a linear least squares target function χ^2, which was minimized using a multi-parameter regression algorithm to determine an optimized solution to the master equation (see SI for further details). We compared the results of our analyses using linear versus cyclical three-state network models, which showed that the cyclical model is favored. We confirmed that our optimized results represent a global minimum of the parameter space by monitoring the increase of the χ^2 error function with variation of the time constant parameters (see Fig. S11 of the SI). The equilibrium populations and kinetic time constants resulting from of our analysis of the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct are summarized in Table 2 and Table 3, respectively, and those of all four ss-dsDNA constructs are presented in Table S1 and Table S6 of the SI.

Table 3. Optimized forward and backward time constants corresponding to the elementary reactive steps obtained for the 3'-Cy3-Cy5-oligo(dT)$_{15}$-ss-dsDNA construct. The time constant parameters are defined in Eq. (9). Error bars were determined from the cross-sections of the error function surfaces $\boldsymbol{\Delta\chi^2/\chi_0^2}$, shown in Fig. S11 of the SI.

Construct	k_{12}^{-1} (ms)	k_{21}^{-1}(ms)	k_{13}^{-1} (ms)	k_{31}^{-1} (ms)	k_{23}^{-1} (ms)	k_{32}^{-1} (ms)
3'-oligo(dT)$_{15}$	0.113 ±0.004	0.025 ±0.0005	104 ±1.85	1.85 ±0.03	4.01 ±0.03	0.325 ±0.03

In Fig. 4*B*, we show the model equilibrium PDF obtained from our optimization analysis of the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct. We have modeled the PDF as a sum of three component Gaussians, with normalized areas equal to the equilibrium populations of the three

macrostates (see Table 2). Thus, macrostate S_1 is the dominant component with p_1^{eq} = 0.81 and mean FRET efficiency $\langle E_{FRET} \rangle_1$ = 0.77 $\pm$ 0.01, macrostate S_2 is a minority component with p_2^{eq} = 0.18 and mean FRET efficiency $\langle E_{FRET} \rangle_2$ = 0.52 $\pm$ 0.07, and macrostate S_3 is a trace component with p_3^{eq} = 0.01 and mean FRET efficiency $\langle E_{FRET} \rangle_3$ = 0.23 $\pm$ 0.07. These optimized Gaussian parameters can be used to estimate the relative free energy minima of the macrostates using the Boltzmann relation [Eq. (2)].

In Fig. 4*D*, we plot the individual macrostate contributions to the free energy surface of the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct, which are shown as dashed purple curves. In Fig. S8 of the SI, we compare the free energy surfaces for all four ss-dsDNA constructs. The sum of the macrostate contributions is shown as a solid blue curve. The three macrostates are represented by parabolas with minimum values G_1^o = 0 (taken as the ground state), G_2^o = 1.11 k_BT and G_3^o = 3.62 k_BT. Here, the superscript 'o' indicates that we have taken the minimum of macrostate S_1 as the reference for all other energies. The parabolic dependence of the free energy with respect to the E_{FRET} parameter is due to our assumption of a Gaussian distribution of end-to-end distances of the oligo(dT)$_n$ template for a given macrostate. The Gaussian distribution describes approximately the relative probability to observe a microscopic chain configuration with an end-to-end distance (and a corresponding E_{FRET} value) that deviates from the mean value, and thus accounts for the entropic contribution to the free energy penalty associated with chain elongation or compression (10). The Gaussian model is, of course, accurate for long ideal polymer chains, and ignores the effects of excluded volume and other internal chain segment interactions. For the oligo(dT)$_n$ template strands, the random walk statistics of chain configurations is additionally influenced by attractive and repulsive interactions between non-adjacent nucleotides and the tendency of adjacent bases to stack locally. Thus, the presence of three distinct macrostates, each with a different mean end-to-end distance, suggests that there exist chain configurations of the oligo(dT)$_n$ templates with characteristically different local flexibilities, which is (presumably) due to the effects of different distributions of stacked bases. We note that in the context of the Gaussian macrostate model, the points at which the component parabolas (dashed purple curves) cross in the free energy surface approximate the entropic contributions to the transition state barriers for macrostate interconversion. However, the true free energy barriers are likely to include enthalpic contributions associated with the energy needed to rearrange the internal constraints imposed by a given

distribution of stacked bases associated with a given macrostate. We next examine the free energy barriers of activation that result from the kinetic analysis of our data.

In Fig. 4*C*, we show the time-dependence of the optimized conditional probabilities $p_{ij}(t)$ where $i, j \in \{1,2,3\}$. These functions illustrate the kinetic behavior of the three-state system. The equilibrium populations are the asymptotic values approached by the conditional probabilities in the limit of long times. For example, the diagonal terms $p_{ii}(t)$ decay from unity to the equilibrium population value p_i^{eq} on the time scale $\left[\sum_{j\neq i=1}^{3} k_{ji}\right]^{-1}$, while off-diagonal terms $p_{ji}(t)$ increase from zero to the equilibrium population value p_i^{eq} on the time scale k_{ji}^{-1}. The relaxation dynamics of all nine conditional probability terms occur primarily on sub-millisecond time scales. The optimized values of the time constants for the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct are presented in Table 3, and the time constants for all four of the ss-dsDNA constructs are listed in Table S6 of the SI. It is interesting to note that the fastest interconversion processes are between macrostates S_1 and S_2 (k_{12}^{-1} = 113 μs and k_{21}^{-1} = 25 μs), while the slowest interconversions are between macrostates S_1 and S_3 (k_{13}^{-1} = 104 ms and k_{31}^{-1} = 1.85 ms) and between macrostates S_2 and S_3 (k_{23}^{-1} = 4.01 ms and k_{32}^{-1} = 325 μs).

From the optimized time constants, we determined the free energies of activation $G_{ij}^{o\ddagger}$ using the Arrhenius relation [Eq. (3)]. In calculating the various values for $G_{ij}^{o\ddagger}$, we have used the maximum rate constant resulting from our analysis ($k_{max} \equiv k_{21}^{-1}$ = 25 μs) as a reference, which we assume corresponds to a barrierless downhill process for the transition $S_2 \rightarrow S_1$ (i.e., $G_{21}^{o\ddagger} = 0$). The remaining activation energies are measured relative to this reference. In Table 4, we list these activation energies for the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct, and we compare the results of all four ss-dsDNA constructs in Table S7 of the SI. We plot these barrier heights as horizontal lines in the free energy diagram shown in Fig. 4*D*. It is interesting to note that the transition barrier corresponding to the uphill reaction $S_1 \rightarrow S_2$ is $G_{12}^{o\ddagger}$ = 1.51 k_BT (green curve) is fully accounted for by the estimated entropic contribution $-TS_{12}^{o\ddagger} \approx$ 1.50 k_BT given by the intersection between the Gaussian components. We thus conclude that the activation barrier for the $S_1 \rightarrow S_2$ transition is purely entropic such that $H_{12}^{o\ddagger} \approx 0$. On the other hand, the transition barriers corresponding to the forward and backward reactions $S_2 \rightleftharpoons S_3$ are $G_{23}^{o\ddagger}$ = 5.08 k_BT and $G_{32}^{o\ddagger}$ = 2.56 k_BT, respectively. The absolute barrier heights corresponding to these values (5.08 + 1.11 = 6.19

k_BT and 2.56 + 3.62 = 6.18 k_BT, respectively), which are indicated on the free energy diagram in Fig. 4*D* (orange curve), exceed the estimated entropic contribution $-TS_{23}^{o\ddagger} \approx 4.65\ k_BT$, suggesting that the forward and backward transitions between macrostates S_2 and S_3 requires $H_{23}^{o\ddagger} \approx H_{32}^{o\ddagger} \cong 1.53\ k_BT$ of internal redistribution energy. Similarly, the forward and backward transition barriers separating macrostates S_1 and S_3 are $G_{13}^{o\ddagger} = 8.34\ k_BT$ and $G_{31}^{o\ddagger} = 4.30\ k_BT$, respectively. The absolute barrier heights corresponding to these values (8.34 + 0 = 8.34 k_BT and 4.30 + 3.62 = 7.92 k_BT, respectively) are plotted on the free energy surface shown in Fig. 4*D* as a red curve, and exceed the estimated entropic contribution $-TS_{13}^{o\ddagger} \approx 6.28\ k_BT$, suggesting that the enthalpic contribution to the activation energies are $H_{13}^{o\ddagger} \approx 2.06\ k_BT$ and $H_{31}^{o\ddagger} \approx 1.64\ k_BT$. We list the estimated enthalpic and entropic contributions to the activation barriers of all four ss-dsDNA constructs in Table S8 of the SI. We note that these values are very similar for the four constructs, and the enthalpic contribution to the forward and backward activation barriers for both the $S_1 \rightleftharpoons S_3$ and the $S_2 \rightleftharpoons S_3$ processes are within the range ~1 – 3 k_BT, which is of comparable magnitude to the enthalpy of base unstacking obtained for TT pairs (1.04 kcal mol^{-1}) in thermal denaturation studies of nicked DNA (43).

Table 4. Optimized free energy minima and transition barriers for the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct. The S_1 macrostate is taken to be the ground state and excited state energies are given by $G_i^o = -k_BTln\left(p_i^{eq}/\sigma_i\sqrt{2\pi}\right) - G_1$ [Eq. (2)] with parameters p_i^{eq} and σ_i listed in Table 2. The activation energies are determined by $G_{ij}^{o\ddagger} = -ln\left(k_{ij}/k_{max}\right)$ [Eq. (3) of the main text, with $k_{max} = k_{21}$ the fastest time constant of the system]. The rate constant parameters are listed in Table S6. Energies are listed in units of k_BT.

Construct	G_1^o	G_2^o	G_3^o	$G_{12}^{o\ddagger}$	$G_{21}^{o\ddagger}$	$G_{13}^{o\ddagger}$	$G_{31}^{o\ddagger}$	$G_{23}^{o\ddagger}$	$G_{32}^{o\ddagger}$
3'-oligo(dT)$_{15}$	0	1.11	3.62	1.52	0	8.34	4.30	5.08	2.56

Table 5. Transition barrier enthalpies and entropies for the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct. The entropic contribution to the transition barrier is defined $-TS_{ij}^{o\ddagger} = G_{ij}^{\text{intersect}}$ with $G_{ij}^{\text{intersect}}$ the free energy at the point of intersection between the parabolas describing macrostate-*i* and macrostate-*j* (dashed purple curves in Fig. 4*D*). The enthalpy of transition is estimated to be $H_{ij}^{o\ddagger} = G_{ij}^{o\ddagger} - G_{ij}^{\text{intersect}}$. Enthalpic and entropic contributions to the transition barriers are given in units of k_BT.

Construct	$-TS_{12}^{o\ddagger}$	$-TS_{13}^{o\ddagger}$	$-TS_{23}^{o\ddagger}$	$H_{12}^{o\ddagger}$	$H_{21}^{o\ddagger}$	$H_{13}^{o\ddagger}$	$H_{31}^{o\ddagger}$	$H_{23}^{o\ddagger}$	$H_{32}^{o\ddagger}$
3'-oligo(dT)$_{15}$	1.50	6.28	4.65	0	0	2.06	1.64	1.54	1.53

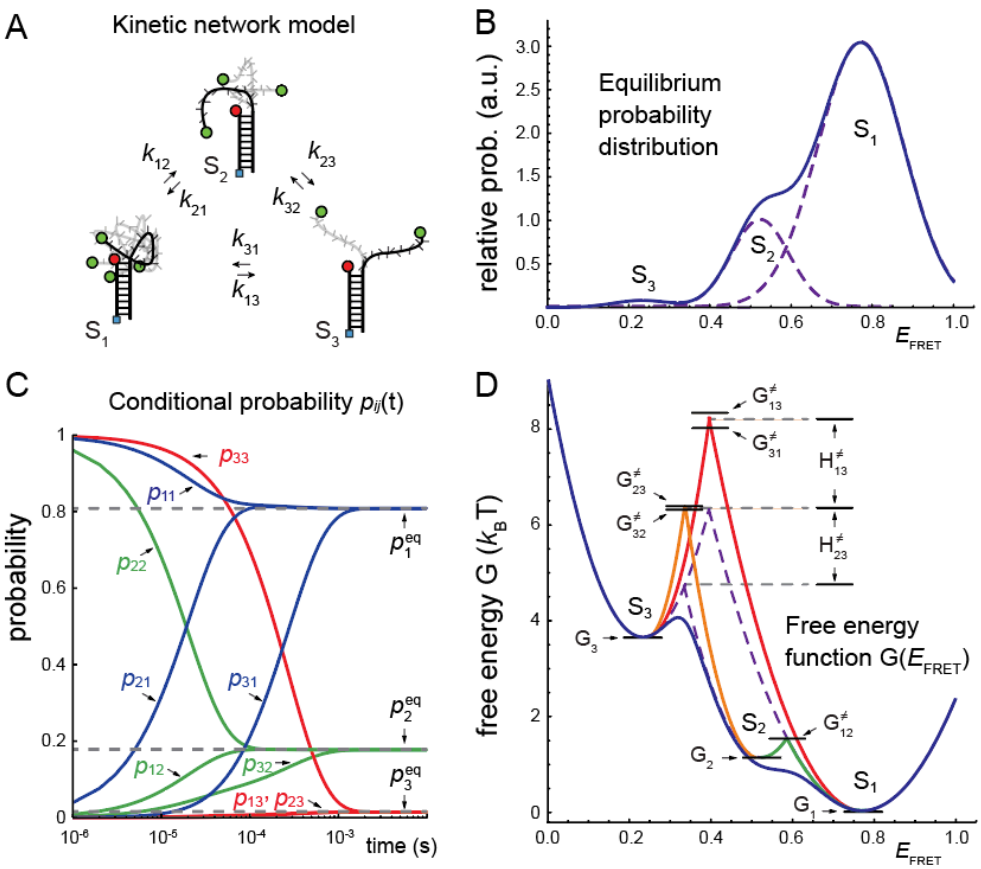

Figure 4. Theoretical modeling results of microsecond-resolved smFRET measurements of the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct using the kinetic master equation. (***A***) Kinetic scheme depicting the elementary reactive steps connecting the three macrostates of the oligo(dT)$_{15}$ template: 'compact' S_1, 'partially-extended' S_2, and 'highly-extended' S_3. (***B***) The normalized probability distribution function (PDF) of E_{FRET} values obtained from an optimized fit to the histogram shown in Fig. 3*A*. The distribution is represented as a sum of the $M = 3$ Gaussian macrostates with areas equal to the equilibrium probabilities p_i^{eq}, with $i \in \{1,2,3\}$. (***C***) The optimized solution to the master equation is the equilibrium PDF, $\boldsymbol{p}^{eq}$, and the time-dependent conditional probabilities, $p_{ij}(t)$, which give the likelihood that the system occupies the *j*th macrostate at time *t* if it previously occupied the *i*th macrostate at time 0, with $i, j \in \{1,2,3\}$. (***D***) The free energy landscape is constructed from the PDF shown in panel *B* according to the Boltzmann distribution [Eq. (2)]. Intersecting dashed purple lines indicate the entropic contributions to the free energy barriers that separate macrostates. Intersecting solid curves indicate the total free energy barriers calculated from the optimized rate constants according to the Arrhenius equation [Eq. (3)]. The optimized values obtained for the equilibrium PDF are given in Table 2, those for the rate

constants are given in Table 3, and the free energy minima and transition barriers are given in Table 4. Estimates of the enthalpic and entropic contributions to the transition barriers are given in Table 5.

5. Conclusions

In this paper we use microsecond-resolved smFRET experiments, recently developed in our laboratory, to begin to characterize conformational fluctuations in the structures of ssDNA at low concentrations in dilute aqueous solution. This is important for studies of protein-DNA interactions because – unlike dsDNA – the interactions stabilizing ssDNA conformations are not cooperative, and therefore conformations that exist at equilibrium, even at trace concentrations, are measurable with our present techniques. These results are likely to be significant from a biological perspective because our approach extends smFRET techniques to the microsecond time regime. By expanding the range of experimentally accessible time scales, it is thus possible to monitor not only the binding and unbinding of protein components of protein-DNA complexes on millisecond time scales and longer, but also to observe directly the interconverting conformational species (on sub-millisecond time scales) of ssDNA that comprise the interaction partners in the initial steps of biologically relevant protein-ssDNA interactions.

As described in the Introduction, and also in previous papers (14, 18, 19, 26), we have been using smFRET methods to study, in particular, the initial steps in the binding, assembly, and translocation of the T4 ssb proteins (gp32), and their cooperatively-bound clusters onto and along ssDNA lattices. Understanding the molecular details of these interactions with both the ssDNA lattices and ss-dsDNA junctions, which are present in the DNA constructs that we have studied, are central to obtaining molecular insight into the regulatory mechanisms of these ssb proteins as they interact with and control the functions of the other components of the bacteriophage T4 replisome – including the helicase and primosome, helicase loader proteins, the DNA polymerases and the processivity clamps and clamp loaders.

In the current paper, we describe microsecond-resolved studies on the ssDNA 'tails' of the ss-dsDNA constructs that we have used previously in our studies of the gp32 binding and assembly system (see Fig. 1). We studied short ss oligo-deoxythymidine [oligo(dT)$_n$] templates of varying length (n = 14 and 15 nts) and polarity (3' versus 5', Table S1 of the SI) because gp32 binding and assembly onto these ss-dsDNA constructs depends sensitively on these factors. As we have shown,

and as one might expect, these differences in the details of the constructs do not significantly change the equilibrium distribution of the template conformations, nor the dynamics of state-to-state interconversion. Nevertheless, it was necessary to investigate all four ss-dsDNA constructs to provide a baseline for the ssb interaction studies that will follow (37). Thus, the present study enables us to re-examine the ssb-ssDNA interaction systems with the ability to determine how ssb protein binding and assembly is directly coupled to the dynamics of the DNA interaction partners.

The oligo(dT)$_n$ templates used in the present study comprised the ssDNA 'tails' of ss-dsDNA constructs, which had been labeled at their ends with a Cy3 donor chromophore and a Cy5 acceptor. We analyzed our data by constructing statistical functions; i.e., the equilibrium probability distribution function of FRET efficiency values and the two-point and four-point time correlation functions. We simulated these functions using a kinetic network model based on a master equation to obtain optimized parameters from which we have reached the following conclusions.

Our results indicate that the oligo(dT)$_n$ templates of all four of the ss-dsDNA constructs used are dynamic systems with three distinct conformational macrostates: a 'compact' form S_1, a 'partially-extended' form S_2, and a 'highly-extended' form S_3 (see Fig. 4*A*). These macrostates are separated in relative stability by just a few thermal energy units: $G_1^o = 0\ k_BT$, $G_2^o = 1.11\ k_BT$, and $G_3^o = 3.62\ k_BT$ in the case of the 3'-Cy3-Cy5-oligo(dT)$_{15}$-ss-dsDNA construct (see Fig. 4*D* and Table 4). The time scales of interconversion between these macrostates span four orders of magnitude, and range from tens-of-microseconds to hundreds-of-milliseconds (see Fig. 4*C* and Table 3). The fastest relaxation time that we observed ($k_{21}^{-1} = 25\ \mu$s) corresponds to the transition from the partially-extended macrostate S_2 to the compact macrostate S_1, which is commensurate with the relaxation time found for ssDNA loop closure in the formation of a DNA hairpin composed of an oligo(dT)$_{16}$ loop (44, 45).

The relative stabilities and relatively long lifetimes of the three macrostates may be related to the energetic cost of unstacking and stacking adjacent bases. From the kinetic rate constants (Table 3) we determined the free energies of activation (Table 4), which we decomposed into entropic and enthalpic contributions (Table 5). The forward and backward enthalpies of activation that separate the highly-extended macrostate S_3 from the partially-extended and compact macrostates, S_2 and S_1, respectively, have values within the range 1.5 to 2 k_BT. If we consider the

barriers for all four of the ss-dsDNA constructs (see Table S8 of the SI), the range of forward and backward enthalpy values increases from 1.2 to 3.1 k_BT. We note that these values are comparable to the enthalpy of base unstacking for a single TT pair (1.04 kcal mol^{-1} = 1.73 k_BT), which was obtained from thermal denaturation studies of the bent-stacked equilibrium in nicked DNA (43). Our estimates of the enthalpic contributions to the activation barriers suggest that only a small number of base stacking interactions are involved in these transformations.

The notion that multiple extended conformations of short oligomeric ssDNA templates exist as stable macrostates at equilibrium differs from our previous assumptions about short ssDNA lattices, in which only a single, relatively compact macrostate has been considered to be the primary conformation (14). The parameters that we have determined to characterize the free energy surface suggest a new model for ssDNA structure may be required (see Fig. 5). In order to gain additional insight into the nature of the extended macrostates, we carried out a coordinate transformation (24) of the free energy surface as a function of the FRET efficiency (shown in Fig. 4*D*) to one as a function of the donor – acceptor separation, which we equate with the end-to-end distance, R_{ee}, of the oligo(dT)$_{15}$ template (Fig. 5*A*). In performing this transformation, we used a value of the Förster transfer distance R_0 = 56 Å obtained by others, which assumes that the relative orientation of the Cy3 donor and Cy5 acceptor, which are attached to the ss-dsDNA constructs using flexible linkers, is random and that the orientation factor κ^2 = 2/3 (31, 32). These assumptions are further validated by our own recent smFRET experiments using polarized light excitation and detection (46). From the transformed free energy surface, we see that the local minima of the three macrostates, S_1 – S_3, may be assigned to mean end-to-end distances $\langle R_{ee} \rangle$ = 46.5, 56.0 and 67.5 Å, respectively.

We next consider how the extended macrostates S_2 and S_3 might play a biological role as intermediates for protein binding mechanisms (47, 48). For example, the T4 ssDNA binding (ssb) protein may be able to access its 7 nt binding site footprint from extended macrostate conformations, S_2 and S_3, more readily than from the compact macrostate S_1. It is known from electron microscopy studies of cooperatively bound ssb nucleoprotein filaments that adjacent nucleotide bases within the complexes are fully unstacked with an average inter-base separation of 4.6 Å (5). This is in contrast to the 3.4 Å separation between stacked bases in B-form duplex DNA. The oligo(dT)$_{14}$ and oligo(dT)$_{15}$ templates examined in this study were chosen specifically

to permit two ssb proteins to bind cooperatively within the final-formed nucleoprotein complexes. In the case of the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct, the contour length of the ssDNA component within a cooperatively-bound (ssb)$_2$-oligo(dT)$_{15}$ filament is expected to be approximately ~15 × 4.6 Å = 69 Å. Given that the mean end-to-end distance corresponding to macrostate S$_3$ (R_{ee} = 67.5 Å) is only slightly smaller than the contour length of the fully unstacked oligo(dT)$_{15}$ template strand, we suggest that the extended macrostates do indeed represent potential loading sites for ssb nucleoprotein filament assembly.

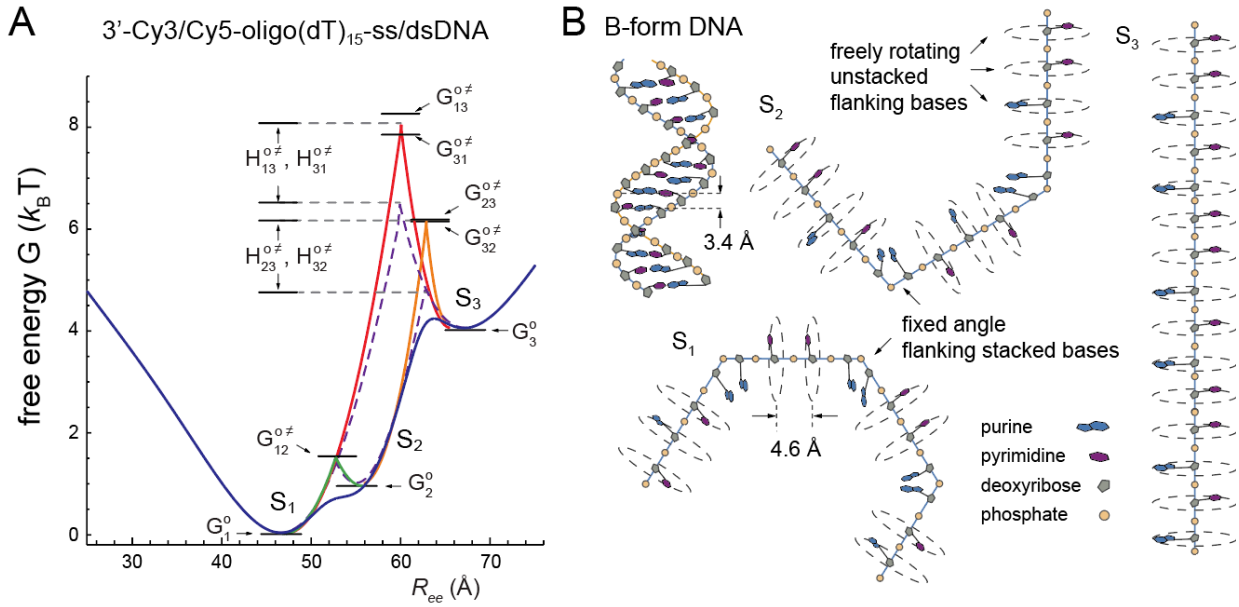

Figure 5. (***A***) Transformed free energy surface as a function of the end-to-end separation, R_{ee}, of the oligo(dT)$_{15}$ template. The transformation is given by $P(R_{ee}) = P(E_{FRET})\, dE_{FRET}/dR_{ee}$ with $R_{ee} = R_0[(1 - E_{FRET})/E_{FRET}]^{\frac{1}{6}}$ and R_0 = 56 Å, and assumes that the relative orientation between the FRET Cy3 donor and Cy5 acceptor is random such that the orientation factor κ^2 = 2/3 (31, 32). (***B***) Schematic microstate configurations of the oligo(dT)$_{15}$ template exemplifying the three macrostates S$_1$ – S$_3$ described by the free energy surface shown in panel (***A***). In B-form DNA (upper left), the mean distance (rise per residue) between adjacent bases is 3.4 Å. In the oligo(dT)$_{15}$ template strands (labeled S$_1$ – S$_3$), adjacent bases are primarily unstacked with a mean inter-base distance of ~4.6 Å. Unstacked bases are stabilized by the gain in entropy associated with the resultant enhanced rotation about the segmental axes. A sequence of unstacked bases forms a contiguous linear segment. The presence of a single stacked base 'defect site,' introduces a 'vertex' with an accompanying change in the bond vector.

In light of our findings, we propose a new model for ssDNA secondary structure, which is depicted schematically in Fig. 5*B*. We posit that at physiological conditions, an isolated ssDNA template exists at thermal equilibrium with a significant fraction of its flanking bases unstacked. In the B-form DNA duplex, the highly ordered secondary structure of the double-helix is stabilized by cooperative short-range attractive interactions (stacking) between flanking bases (depicted as blue purines and red pyrimidines) and the network of Watson-Crick hydrogen bonds between opposing bases of the complementary single-strands. The duplex is marginally destabilized by the Coulomb repulsions between negatively charged phosphates (shown as yellow circles) of the sugar-phosphate backbones, which are partially screened by the condensation layer of positively charged counterions in the immediate aqueous surroundings.

We conjecture that adjacent stacked bases within an isolated segment of ssDNA, which lack the stabilizing influence of the complementary strand of duplex DNA, are only marginally stabilized with respect to thermal fluctuations that disrupt the short-range (solvophobic) attraction between base surfaces (49). Unstacked bases within an isolated segment of ssDNA are energetically favored by the Coulombic repulsion between adjacent phosphates and the gain in entropy associated with enhanced rotational freedom about the segmental axes. A sequence of metastable unstacked bases forms an essentially linear contiguous segment, which may be disrupted by the spontaneous formation of a stacking interaction between two adjacent bases along the ssDNA chain. The presence of such a vicinal stacked pair of bases introduces a 'defect site,' which then induces a change in the bond vector of two adjoining segments, each composed of unstacked bases. In Fig. 5*B*, we depict three examples of such microstate ssDNA conformations, which may each be assigned to one of the three macrostates S_1 – S_3, as shown. Each macrostate can be viewed as a stable family of microstates that contains a unique, dynamically equilibrated distribution of stacked base 'defect sites.' The principal difference between the underlying distributions of the three macrostates is the mean number of stacked base defect sites, whose presence effectively disrupts the linearity of the oligo$(dT)_{15}$ template, and thereby reduces the end-to-end distance. The mean number of defect sites is maximized for the most stable macrostate S_1 (depicted in the figure having three defect sites). The ground state S_1 minimum represents the end-to-end distance ($\langle R_{ee} \rangle$ = 46.5) corresponding to balance between stabilizing and de-stabilizing (entropic and enthalpic) interactions, which arise in the presence of the dynamically interconverting stacked base defect sites. We further note that in the context of this model,

conformational transformations between any two of the three macrostates, $S_1 – S_3$, correspond to the loss or gain of a single stacked base defect site, which is fully consistent with the values that we have determined for the relative stabilities and transition enthalpy barriers of the macrostates.

The above model for ssDNA secondary structure can provide useful insights for the interpretation of future microsecond-resolved smFRET studies of the T4 bacteriophage ssb nucleoprotein filament assembly mechanism using these same ss-dsDNA constructs. The initial steps of ssb protein-filament assembly involve the association of a ssb (gp32) monomer with an exposed ssDNA template binding site (7 nts), which may form on sub-millisecond time scales as a contiguous segment of unstacked flanking bases. Additional functional steps in which ssb proteins can assemble into cooperatively-bound clusters, slide along a ssDNA template strand and disassemble, will be mediated by the nucleation and dissolution of flanking stacked base defect sites, as described by our model. We anticipate that the model can be further tested by performing microsecond-resolved smFRET experiments on ss-dsDNA constructs with ssDNA sequences of purine or of mixed base sequences.

Acknowledgements

A.H.M. acknowledges useful discussions with Prof. Marina Guenza. The authors are also grateful to our laboratory colleagues in the Marcus and von Hippel groups for many useful discussions. This work was supported by grants from the National Institutes of Health General Medical Sciences (GM-15792 to A.H.M. and P.H.v.H.). B.I. and C.S.A. were supported as predoctoral trainees by an NIH-NIGMS Institutional Research Service Award in Molecular Biology and Biophysics (GM-07759). P.H.v.H. is an American Cancer Society Research Professor of Chemistry.

References

1. K. Luger, M. L. Dechassa, and D. J. Tremethick, New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature* **13**, 436-447 (2012).

2. M. Spies, and T. Ha, Inching over hurdles: How DNA helicases move on crowded lattices. *Cell Cycle* **9**, 1742-1749 (2010).

3. A. G. Kozlov, M. K. Shinn, and T. M. Lohman, Regulation of nearest-neighbor cooperative binding of E. coli SSB protein to DNA. *Biophys. J.* **117**, 2120-2140 (2019).

4. P. H. von Hippel, and A. H. Marcus, The many roles of binding cooperativity in the control of DNA replication: Invited New and Notable Commentary related to Kozlov, Shinn and Lohman, Biophys. J., 2019, 117, 2120-40. *Biophys. J.* **117**, 2143-2146 (2019).

5. H. Delius, N. J. Mantell, and B. Alberts, Characterization by electron microscopy of the complex formed between T4 bacteriophage gene 32-protein and DNA. *J. Mol. Biol.* **67**, 341-350 (1972).

6. M. Duane, *Molecular Biophysics: Structures in Motion* (Oxford University Press, Oxford, 1993).

7. M. Muthukumar, 50th anniversary perspective: A perspective on polyelectrolyte solutions. *Macromolecules* **50**, 9528-9560 (2018).

8. D. Laage, T. Elsaesser, and J. T. Hynes, Water dynamics in the hydration shells of biomolecules. *Chem. Revs.* **117**, 10694-10725 (2017).

9. D. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*, Cambridge Molecular Science (Cambridge University Press, Cambridge, 2004).

10. M. Rubinstein, and R. H. Colby, *Polymer Physics* (Oxford University Press, Oxford, 2003).

11. F. Noé, D. Krachtus, J. C. Smith, and S. Fischer, Transition networks for the comprehensive characterization of complex conformational change in proteins. *J. Chem. Theory Comput.* **2**, 840-857 (2006).

12. H.-S. Chung, I. V. Gopich, K. McHale, T. Cellmer, J. M. Louis, and W. A. Eaton, Extracting rate coefficiencts from single-molecule photon trajectories and FRET efficiency histograms for a fast-folding protein. *J. Phys. Chem. A* **115**, 3642-3656 (2011).

13. E. R. Beyerle, M. Dinpahjooh, H. Ji, P. H. von Hippel, A. H. Marcus, and M. G. Guenza, Dinucleotides as simple models of the base stacking-unstacking component of DNA 'breathing' mechanisms. *Nucleic Acids Res* **49**, 1872-1885 (2021).

14. C. Phelps, B. Israels, D. Jose, M. C. Marsh, P. H. von Hippel, and A. H. Marcus, Using microsecond single-molecule FRET to determine the assembly pathways of T4 ssDNA binding protein onto model DNA replication forks. *Proc. Nat. Acad. Sci. Plus USA* **114**, E3612-E3621 (2017).

15. W. Lee, J. P. Gillies, D. Jose, P. H. von Hippel, and A. H. Marcus, Single-molecule FRET studies of cooperative and non-cooperative binding kinetics of the bacteriophage T4 single-stranded DNA binding protein (gp32) to ssDNA lattices at replication fork junctions. *Nucleic Acids Res* **44**, 10691-10710 (2016).

16. M. C. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha, Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.* **86**, 2530-2537 (2004).

17. T. Ha, T. M. Lohman, and A. G. Kozlov, Single-molecule views of protein movement on single-stranded DNA. *Ann. Rev. Biophys.* **41**, 295-319 (2012).

18. C. Phelps, W. Lee, D. Jose, P. H. von Hippel, and A. H. Marcus, Single-molecule FRET and linear dichroism studies of DNA 'breathing' and helicase binding at replication fork junctions. *Proc Natl Acad Sci U S A* **110**, 17320 – 17325 (2013).

19. C. Phelps, B. Israels, D. Jose, M. C. Marsh, P. H. von Hippel, and A. H. Marcus, Using multi-order time correlation functions (TCFs) to elucidate biomolecular reaction pathways from microsecond single-molecule fluorescence experiments. *J. Phys. Chem. B* **120**, 13003-13016 (2016).

20. A. Tamimi, T. Landes, J. Lavoie, M. G. Raymer, and A. H. Marcus, Fluorescence-detected Fourier transform electronic spectroscopy by phase-tagged photon counting. *Opt. Express* **28**, 25194-25214 (2020).

21. M. A. Berg, and J. R. Darvin, Measuring a hidden coordinate: Rate-exchange kinetics from 3D correlation functions. *J. Chem. Phys.* **145**, 054119 (2016).

22. M. A. Berg, and H. Kaur, Nonparametric analysis of nonexponential and multidimensional kinetics. I. Quantifying rate dispersion, rate heterogeneity, and exchange dynamics. *J. Chem. Phys.* **146**, 054104 (2017).

23. M. Kryvohuz, and S. Mukamel, Nonlinear response theory in chemical kinetics. *J. Chem. Phys.* **140**, 034111 (2014).

24. G. Beddard, *Applying Maths in the Chemical and Biological Sciences: An Example-Based Approach* (Oxford University Press, Oxford, 2009).

25. L. E. Reichl, *A Modern Course in Statistical Physics* (John Wiley & Sons, New York, 1998).

26. W. Lee, D. Jose, C. Phelps, A. H. Marcus, and P. H. von Hippel, A single-molecule view of the assembly pathway, subunit stoichiometry and unwinding activity of the bacteriophage T4 primosome (helicase-primase) complex. *Biochemistry* **52**, 3157-3170 (2013).

27. C. Joo, S. A. McKinney, M. Nakamura, I. Rasnick, S. Myong, and T. Ha, Real-time observation of RecA filament dynamics with single monomer resolution. *Cell* **126**, 515-527 (2006).

28. I. Rasnik, S. A. McKinney, and T. Ha, Nonblinking and long-lasting single-molecule fluorescence imaging. *Nature Meths.* **3**, 891-893 (2006).

29. S. D. Chandradoss, A. C. Haagsma, Y. K. Lee, J.-H. Hwang, J.-M. Nam, and C. Joo, Surface passivation for single-molecule protein studies. *J. Vis. Exper.* **86**, e50549 (2014).

30. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy* (Kluwer Acad. / Plenum, New York, 1999).

31. R. Screenivasan, I. A. Shkel, M. Chhabra, A. Drennan, S. Heitkamp, H.-C. Wang, M. A. Sridevi, D. Plaskon, C. McNerney, K. Callies, C. K. Cimperman, and M. T. Record Jr., Fluorescence-detected conformational changes in duplex DNA in open complex formation by *Escherichia coli* RNA polymerase: Upstream wrapping and downstream bending

precede clamp opening and insertion of the downstream duplex. *Biochemistry* **59**, 1565-1581 (2020).

32. R. Screenivasan, I., M. Chhabra, R. Saecker, E. Lingeman, M. Poulos, D. McCaslin, M. W. Capp, I. Artsimovitch, and M. T. Record Jr., Flourescence resonance energy transfer characterization of DNA wrapping in closed and open Escherichia coli RNA polymerase - λPR promoter complexes. *Biochemistry* **55**, 2174-2185 (2016).

33. D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, 1987).

34. M. B. Zimmt, K. A. Peterson, and M. D. Fayer, Short polymer chain statistics and the relationship to end to end electronic excitation transport: Random walks with variable step lengths. *Macromolecules* **21**, 1145-1154 (1988).

35. D. T. Gillespie, Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **58**, 35-55 (2007).

36. D. Glasser, F. J. M. Horn, and R. Meidan, Properties of certain zero column-sum matrices with applications to the optimization of chemical reactors. *J. Math. Anal. Appl.* **73**, 315-337 (1980).

37. B. Israels, C. Albrecht, A. Dang, M. Barney, P. H. von Hippel, and A. H. Marcus, The effects of ssDNA length and polarity on the assembly mechanism of gp32 dimers onto ssDNA templates near ss-dsDNA junctions. *unpublished results* (2021).

38. N. Kumari, M. A. Cuiba, and M. Levitus, Photophysical properties of the hemicyanine Dy-630 and its potential as a single-molecule fluorescent probe for biophysical applications. *Methods Appl. Fluoresc.* **8**, 015004 (2020).

39. M. E. Sanborn, B. K. Connolly, K. Gurunathan, and M. Levitus, Fluorescence properties and photophysics of the sulgoindocyanine Cy3 linked covalently to DNA. *J. Phys. Chem. B* **111**, 11064-11074 (2007).

40. K. Y. W. Jia, A. Xia, S. Li, F. Gong, and G. Yang, Characterization of photoinduced isomerization and intersystem crossing of the cyanine dye Cy3. *J. Phys. Chem. A* **111**, 1593-1597 (2007).

41. M. A. Ciuba, and M. Levitus, Manganese-induced triplet blinking and photobleaching of single molecule cyanine dyes. *ChemPhysChem* **14**, 3495-3502 (2013).

42. V. A. Kuzmin, and A. P. Darmanyan, Study of sterically hindered short-lived isomers of polymethine dyes by laser photolysis. *Chem. Phys. Lett.* **54**, 159-163 (1978).

43. E. Protozanova, P. Yakovchuk, and M. D. Frank-Kamenetskii, Stacked-unstacked equilibrium at the nick site of DNA. *J. Mol. Biol.* **342**, 775-785 (2004).

44. N. L. Goddard, G. Bonnet, O. Krichevsky, and A. Libchaber, Sequence dependent rigidity of single stranded DNA. *Phys. Rev. Lett.* **85**, 2400-2403 (2000).

45. G. Bonnet, O. Krichevsky, and A. Libchaber, Kinetics of conformational fluctuations in DNA hairpin-loops. *Proc. Nat. Acad. Sci.* **95**, 8602-8606 (1998).

46. C. S. Albrecht, J. Maurer, and A. H. Marcus, Determination of the orientation factor in single-molecule FRET experiments by polarization-sweep photon correlation microscopy. *unpublished results* (2021).

47. P. H. von Hippel, "On the molecular bases of the speficity of interaction of transcriptional proteins with genome DNA" in Biological Regulation and Development, R. F. Goldberger, Ed. (Plenum Publishing Corp., New York, 1979), vol. 1, pp. 279-347.

48. P. H. von Hippel, N. P. Johnson, and A. H. Marcus, 50 years of DNA 'breathing': Reflections on old and new approaches. *Biopolymers* **99**, 923-954 (2013).

49. M. J. Lowe, and J. A. Schellman, Solvent effects on dinucleotide conformation. *J. Mol. Biol.* **65**, 91-109 (1972).

Supporting Information: Sub-millisecond conformational transitions of short single-stranded DNA lattices by photon correlation single-molecule FRET

Brett Israels,[1,2] Claire S. Albrecht,[3] Anson Dang,[1] Megan Barney,[1] Peter H. von Hippel,[1] and Andrew H. Marcus[1,2,3*]

[1.] Department of Chemistry and Biochemistry, Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA

[2.] Department of Chemistry and Biochemistry, Center for Optical, Molecular and Quantum Science, University of Oregon, Eugene, OR 97403, USA

[3.] Department of Physics, Center for Optical, Molecular and Quantum Science, Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA

[*]e-mail: ahmarcus@uoregon.edu

1. Materials and Methods

Sample Preparation. We studied four single-stranded (ss) – double-stranded (ds) DNA constructs with a ss oligo(dT)$_n$ overhanging 'tail' region, which varied in length (n = 14 and 15) and polarity (3' and 5', see Fig. 1 and Table 1 of the main text). DNA samples were purchased at nanomole quantities as dehydrated solids composed of individual single-strands (Integrated DNA Technologies, *IDT*, Coralville, IA, USA). The samples were rehydrated in an aqueous buffer composed of 100 mM NaCl, 6 mM $MgCl_2$, and 10 mM Tris (pH 8.0) to reach a concentration of 100 μM. The ssDNA solutions were annealed by first diluting aliquots of the 100 μM Cy3-containing ('template') strand in buffer to reach 150 nM concentration, and the complementary 100 μM Biotin/Cy5-containing ('primer') strand to reach 100 nM concentration. The two solutions were combined to achieve a 2:3 (p:t) number ratio to ensure each primer strand has an attached template. The combined solution was heated to 94°C for 2 minutes and allowed to cool gradually to room temperature 22°C. The annealed ss-dsDNA constructs were stored at 4°C between experiments, or alternatively frozen at –4°C during extended periods. The annealed ss-dsDNA solution was diluted 1000-fold before it was introduced into the microscope sample chamber, as described below.

Microfluidic Sample Chamber. Sample chambers were constructed from microscope slides, which were chemically modified using the procedure described by Chandradoss *et al*. (see Fig. S1*A*) (1). Sample chambers were thoroughly cleaned with detergent (Alconox), water, acetone, KOH, acidic piranha-solutions (Sulfuric Acid + Hydrogen Peroxide), and methanol. The piranha solution served to clean the surface of the slide and to reduce the silicon-oxide to Si-OH. The reduced fused-silica quartz surface was treated with a solution of 3-aminopropyl-trimethoxy silane + acetic acid in methanol, and subsequently with poly(ethylene glycol)-succinimidyl valerate (PEG-SVA) (Laysan Bio, Inc.) with a molecular weight of 5000 Da in 100 mM sodium-bicarbonate solution to passivate the surface against interactions with proteins and to prevent non-specific binding of DNA. A small fraction of the PEG that coated the microscope slides were labeled with biotin. Neutravidin (Thermo Fisher Scientific), a de-glycosylated avidin protein tetramer with strong binding affinity to biotin (2), was used as a linker to bind the biotin attached to the slide to the biotin-covalently bound to the ends of the ss-dsDNA constructs.

Oxygen Scavenging and Triplet State Quenching Reagents. An oxygen scavenging solution consisting of glucose oxidase, catalase and glucose were used to sequester free oxygen in the solution. The solution was introduced into the sample cell before smFRET data were acquired. A series of chemical reactions results in the net loss of oxygen from the buffer (3). In the absence of oxygen scavenging buffer, dissolved oxygen in the solution will irreversibly react with the excited-state fluorophores leading to photobleaching of the sample. In addition, the triplet state quencher Trolox (Sigma Aldrich) was used to prevent 'photo-blinking,' which is a process in which the chromophore probes reversibly interconvert between excited singlet and triplet states (4). Because the lifetimes of the excited triplet states are on the order of milliseconds, the resulting slow recovery of the singlet ground state leads to a dramatic decrease in the fluorescence signal rate.

Microsecond-Resolved Single-Molecule (sm) FRET. We performed microsecond-resolved smFRET experiments on the ss-dsDNA constructs using instrumentation and procedures described previously (5, 6). In Fig. S1*B*, we show a schematic of the instrumental setup. The microfluidic sample chamber was positioned using a computer-controlled translation stage (NPS-XY-100A with NPS3330 controller, Queensgate, UK), which was mounted to an inverted fluorescence microscope (Nikon 2000 TE Eclipse). The sample was illuminated using a continuous wave 532 nm laser (Coherent Compass) in a prism-based total internal conversion fluorescence (TIRF) geometry. The incident laser was focused to a 50 μm diameter spot on the sample using a 100 mm

plano-convex lens, and the ensuing fluorescence was collected using a 100× oil-immersion objective (Nikon CF160, Plan Apochromat, N.A. 1.4, w.d. 130 μm). The fluorescence from a single molecule was transmitted through a 100 μm pinhole (ThorLabs), which was positioned at the downstream image plane of the 200 mm microscope lens tube. The transmitted signal was optimized by scanning the XY-stage with sub-nanometer precision. The fluorescence was collimated using a 75 mm biconvex lens, which was placed one focal length downstream from the pinhole. The fluorescence was directed toward a 635 nm dichroic beam splitter (FF640-FDi01-25x36, Semrock, USA), and the separated donor Cy3 and acceptor Cy5 fluorescence channels were focused onto avalanche photodiodes (APDs, SPCM AQR-13, Perkin-Elmer Optoelectronics, Singapore) using long working distance objective lenses (SLCPlanFl, 40x, N.A. 0.55, w.d. 2.6 mm, Olympus, Hauppauge, NY). To ensure that the molecules we studied were localized within a uniform and uncrowded region of the sample chamber, we used an electron multiplying charge-coupled device camera (emCCD, iXon DV897-BB, 512 × 512 pixels, Andor Technologies, Belfast, Northern Ireland) to acquire wide-field images at 30 ms resolution.

Data Acquisition. The laser power at the sample was adjusted to 10 mW, which resulted typically in a combined donor-acceptor fluorescence photon count rate of ~ 15,000 – 20,000 cps. The signal 'background rate' of ~ 2,000 cps was determined by scanning the XY-stage position away from the center of a molecule. Single-photon detection resulted in the emission of TTL voltage pulses from the APDs, which were sent to a custom built 64 MHz digital counter. The donor Cy3 and acceptor Cy5 signals were each composed of a stream of temporally separated TTL pulses, which were 'time-stamped' at a resolution of 0.1 μs using a 10 MHz digital counter and stored on a computer hard disk. A typical single-molecule data run was recorded for a thirty-second duration. In post-data acquisition, we determined the time-dependent *D* (*A*) signal rates according to $I_{D(A)}(t) = N_{D(A)}(t)/T_w$, where $N_{D(A)}$ was the number of *D* (*A*) photons detected during an adjustable dwell period T_w ($\geq 10\ \mu$sec) at time t. Multiple single-molecule data sets were recorded, each for a thirty-second duration. These signal trajectories were initially screened to reject cases that contained photobleaching events or multiple *D* chromophores. From the time-dependent signal rates we determined the FRET efficiency according to: $E_{FRET}(t) = I_A(t)/[I_D(t) + I_A(t)]$. An example trajectory for the 3'-Cy3/Cy5-oligo$(dT)_{15}$-ss-dsDNA construct is shown in Fig. S2 for an integration window of T_w = 1 ms.

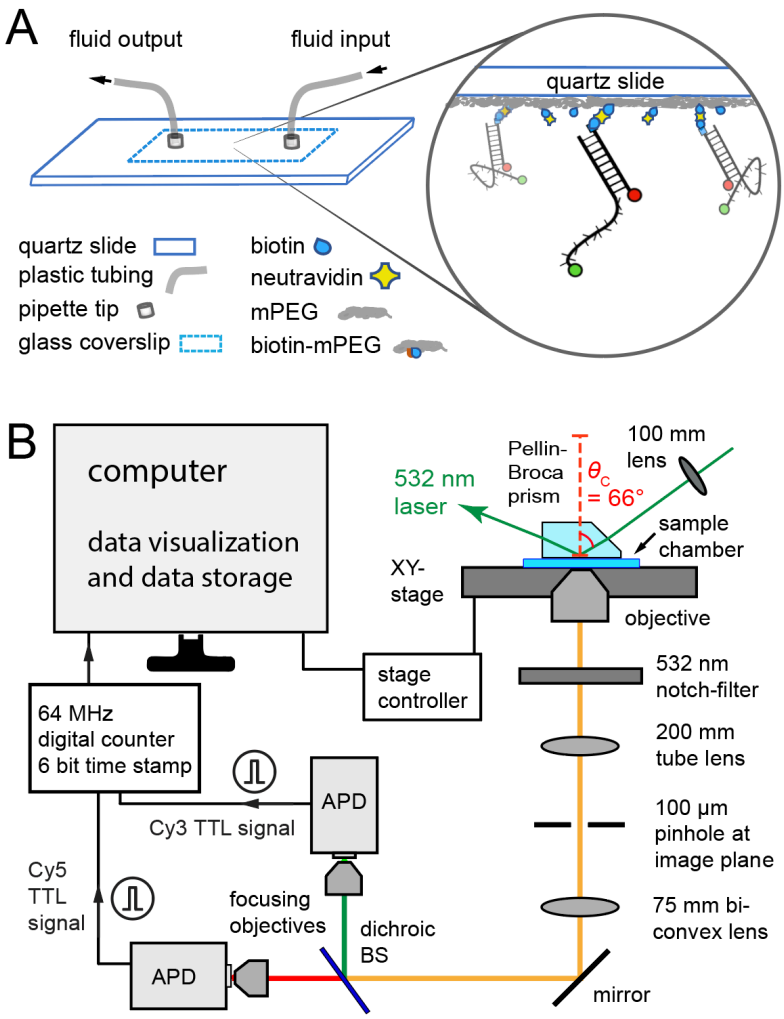

Figure S1. Experimental instrumentation used for microsecond-resolved single-molecule Förster resonance energy transfer (smFRET) experiments described the text. (***A***) Microfluidic sample chambers were constructed from fused-silica quartz microscope slides as described above. (***B***) The experimental setup is a modified version of the one described by Phelps *et al*. (5). APD = avalanche photodiode, BS = beam splitter, TTL = transistor-transistor logic, θ_c = critical angle.

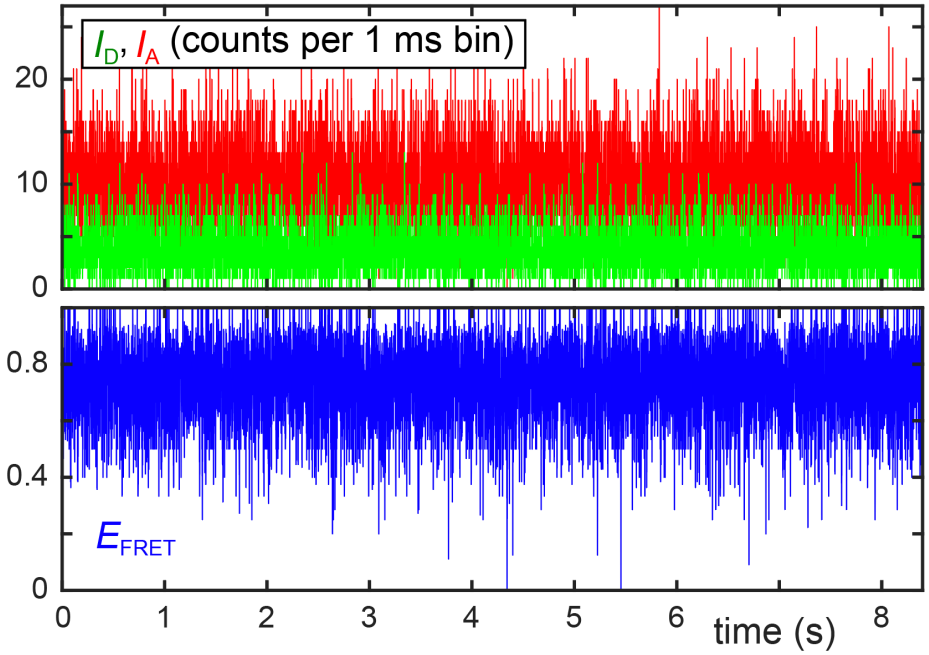

Figure S2. Example smFRET trajectory for the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct. The donor *D* (Cy3, green) and acceptor *A* (Cy5, red) fluorescence intensities, $I_{D(A)}(t) = N_{D(A)}(t)/T_w$, are shown in the top panel with bin resolution T_w = 1 ms. The FRET efficiency $E_{FRET}(t) = I_A(t)/[I_D(t) + I_A(t)]$, which is calculated from the $I_{D(A)}$ signals, is shown in the bottom panel (blue). The full data set was recorded for a duration of 30 seconds, although the trajectory is plotted here over a period of 8.5 seconds.

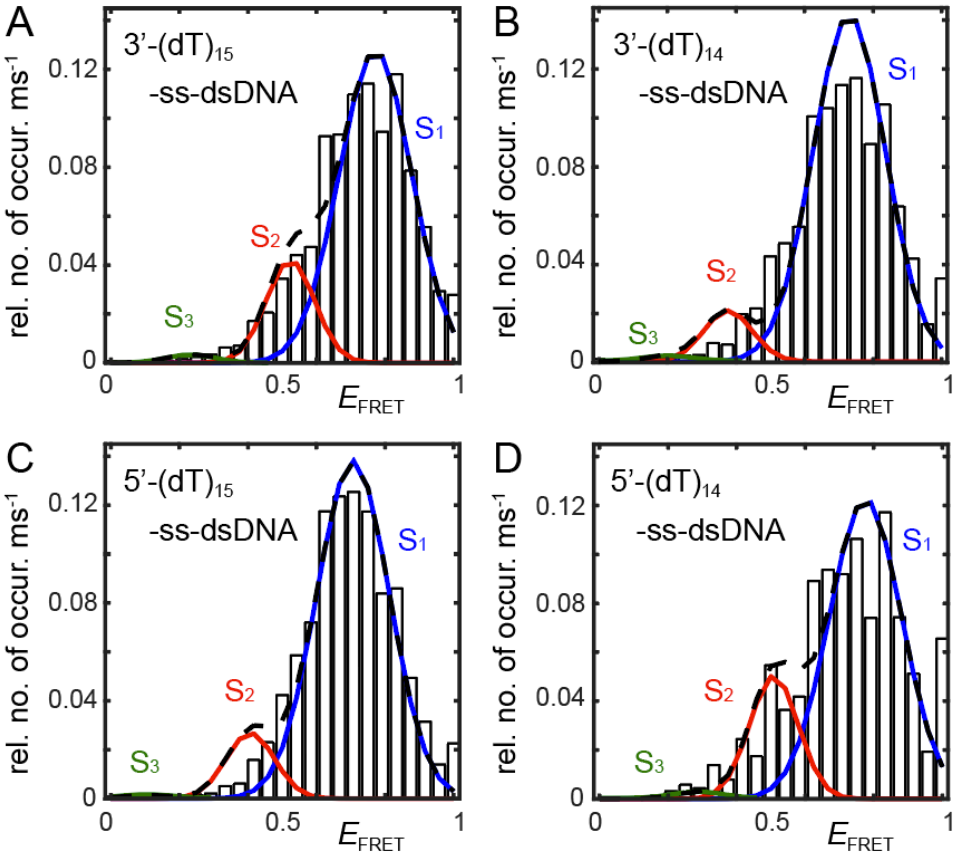

Figure S3. Experimental histograms and PDFs for the four ss-dsDNA constructs. The blue, red, and green Gaussian curves denote the compact (S_1), partially extended (S_2), and highly extended (S_3) macrostate conformations, respectively. The optimized parameters obtained from our analysis of each of these data sets using the three-state master equation are listed in Table S1.

Table S1. Optimized parameters corresponding to the FRET probability distribution functions (PDFs) obtained for the four ss-dsDNA constructs studied in this work. The parameters are defined in the main text in and around Eq. (2), with $\boldsymbol{A_i = p_i^{eq}/\sigma_i\sqrt{2\pi}}$. Error bars for the mean FRET efficiency values are on the order of the last significant figure, and are listed for the 3'-oligo(dT)$_{15}$ construct in Table S11.

Construct	A_1	$\langle E_{FRET}\rangle_1$	σ_1	P_1^{eq}	A_2	$\langle E_{FRET}\rangle_2$	σ_2	P_2^{eq}	A_3	$\langle E_{FRET}\rangle_3$	σ_3	P_3^{eq}
3'-oligo(dT)$_{15}$	3.04	0.772	0.106	0.808	1.001	0.524	0.071	0.177	0.081	0.233	0.071	0.014
3'-oligo(dT)$_{14}$	3.39	0.730	0.106	0.902	0.501	0.376	0.071	0.089	0.051	0.200	0.071	0.009
5'-oligo(dT)$_{15}$	3.31	0.707	0.106	0.880	0.644	0.407	0.071	0.114	0.033	0.104	0.071	0.006
5'-oligo(dT)$_{14}$	2.92	0.776	0.106	0.776	1.202	0.509	0.071	0.213	0.064	0.287	0.071	0.011

Numerical Calculations of Two-Point and Four-Point Time Correlation Functions (TCFs). We calculated the two-point and four-point time correlation function (TCFs) for each single molecule data set. A typical microsecond-resolved single-molecule photon data stream contained ~30 million time points with a mean signal flux ~20,000 s^{-1}. We compared the TCFs constructed from individual molecules for consistency before averaging them together to obtain the final result.

If the photon data stream is binned over a sufficiently long window T_w (typically ~1 ms), the signal intensities, $I_{D(A)}(t) = N_{D(A)}(t)/T_w$ (where $N_{D(A)}$ is the number of D (A) photons detected during the period T_w) are vary continuously with time. In this case, the Wiener-Khinchin theorem can be applied to compute the two-point TCF according to:

$$\bar{C}_E^{(2)}(\tau) = \frac{1}{N_{\text{binned}}^2}\mathcal{F}^{-1}\{|\mathcal{F}[\delta E_{FRET}(\text{t})]|^2\} \tag{S1}$$

where N_{binned}^2 is the number of binned time points over which the FRET efficiency signal $\delta E_{FRET}(\text{t}) = E_{FRET}(\text{t}) - \langle E_{FRET}\rangle$, with $E_{FRET}(t) = I_A(t)/[I_D(t) + I_A(t)]$, is monitored and $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent, respectively, the forward and inverse fast-Fourier-transform (FFT).

Because our microsecond-resolved smFRET signal is sparsely sampled at a much higher frequency than the mean signal flux, the data stream is non-continuous and we cannot use Eq. (S1) to calculate the two-point TCFs. Instead, we implemented the following pair-wise point-by-point multiplication scheme:

$$\bar{C}_E^{(2)}(\tau) = \frac{1}{N_{2P}(\tau)}\sum_{i=0}^{N_{\text{tot}}-1}[\delta E_{FRET}(t_i)\delta E_{FRET}(t_i+\tau)] \tag{S2}$$

In Eq. (S2), the two-point TCF is a rolling time-average of the product of two successive measurements, $\delta E_{FRET}(t_i)\delta E_{FRET}(t_i+\tau)$, which are spaced apart by the interval τ. Here, t_i is the measurement time index, N_{tot} ($\approx$ 30 million) is the total number of evenly spaced data points, and $N_{2P}(\tau)$ is the number of non-zero terms that contribute to the sum. An important consequence of

the sparse data sampling of these measurements is that the number $N_{2P}(\tau)$ depends non-uniformly on the interval τ. By accounting for the precise number of non-zero terms that contribute to the sum, it is possible to determine accurately the amplitude and shape of the two-point TCF. We have compared the results of Eq. (S2) to those using the Wiener-Khinchin formula [Eq. (S1)] for cases in which the signal varies continuously in time. We note that in this limit the two methods provide equivalent results, as expected.

To determine the four-point TCFs, we implemented a point-by-point multiplication scheme similar to the one we used for the two-point TCFs:

$$\bar{C}_E^{(4)}(\tau_1,\tau_2,\tau_3) \tag{S3}$$
$$= \frac{1}{N_{4P}(\tau_1,\tau_2,\tau_3)} \sum_{i=0}^{N_{\text{tot}}-1} [\delta E_{FRET}(t_i)\delta E_{FRET}(t_i+\tau_1)\delta E_{FRET}(t_i+\tau_1+\tau_2)\delta E_{FRET}(t_i+\tau_1+\tau_2+\tau_3)]$$

In Eq. (S3), $N_{4P}(\tau_1,\tau_2,\tau_3)$ is the number of non-zero four-point product terms that contribute to the sum, which are spaced apart by the intervals τ_1, τ_2 and τ_3. For all calculations of the four-point TCFs presented in this work, we have set the intermediate time interval $\tau_2 = 0$.

The experimental two-point and four-point TCFs do not fully decay to zero at long times, but rather decay to non-zero asymptotic values. To account for this, we included an empirically-determined constant offset in all of our decay models. For the two-point TCFs, we used the model $\bar{C}^{(2)}(\tau) = \mathcal{A}_2 e^{\lambda_2\tau} + \mathcal{A}_3 e^{\lambda_3\tau} + B$, and for the four-point TCFs, we used the model $\bar{C}^{(4)}(\tau_1,\tau_2 = 0,\tau_3) = \mathcal{A}_{22} e^{\lambda_2(\tau_1+\tau_3)} + \mathcal{A}_{23} e^{\lambda_2\tau_1+\lambda_3\tau_3} + \mathcal{A}_{32} e^{\lambda_3\tau_1+\lambda_2\tau_3} + \mathcal{A}_{33} e^{\lambda_3(\tau_1+\tau_3)} + D$. The values that we used for the constant offsets of the two-point and four-point TCFs (*B* and *D*, respectively) were determined from the means of the last 3 calculated data points of each surface, corresponding to times at which the decays had reached their asymptotic values.

Instrument Response Function. We performed control measurements on a concentrated solution of Rhodamine 6G (Rh6G), which fluoresces over a broad spectral range that spans both the donor and acceptor emission channels. Successive photon emission events from the concentrated Rh6G solution were uncorrelated on the 0.1 μs resolution of the instrument. The TCF of these data were

used to construct the instrument response function (IRF), which has full-width-at-half-maximum ~ 1 μs. We used the IRF in a deconvolution procedure (7) to accurately determine the fastest time scale components (microseconds) of the TCFs that we determined from our smFRET measurements of the ss-dsDNA constructs (see Fig. S4 and Table S2).

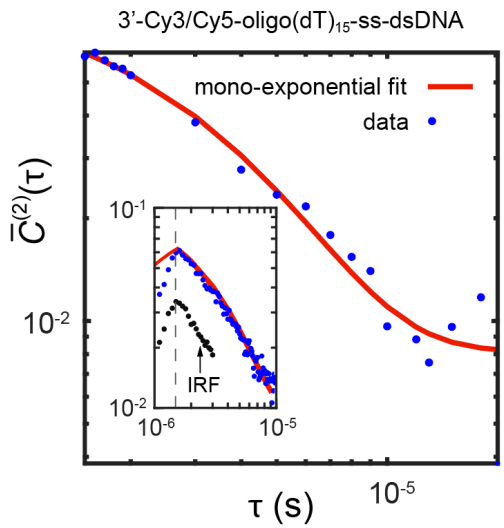

Figure S4. Two-point TCF of the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct is shown over the range of time scales: $1.5 \times 10^{-6} - 2 \times 10^{-5}$ s. The first 20 μs of the autocorrelation function was fit to a model mono-exponential decay using the equation $C^{(2)} = Ae^{-t/\tau_0} + B$, where B is a constant y-axis offset. The inset shows the measured instrument response function used to perform the short-time fits. All four of the ss-dsDNA constructs exhibited a single exponential decay constant ~3 μs (see Table S2), which we interpret to be an effect of the Cy3 excited state photoisomerization process.

Table S2. Decay constants determined from mono-exponential fits to the two-point TCFs over the range of time scales: $10^{-6} - 2 \times 10^{-5}$ s. Each of the four ss-dsDNA constructs exhibited a decay with time constant $\boldsymbol{\tau_0} \approx 3$ μs, as shown for the case of the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct in Fig. S4.

Construct	Time Constant τ_0 (μs)
3'-oligo(dT)$_{15}$	2.9 ± 0.45
3'-oligo(dT)$_{14}$	2.7 ± 0.35
5'-oligo(dT)$_{15}$	3.2 ± 0.51
5'-oligo(dT)$_{14}$	3.0 ± 0.50

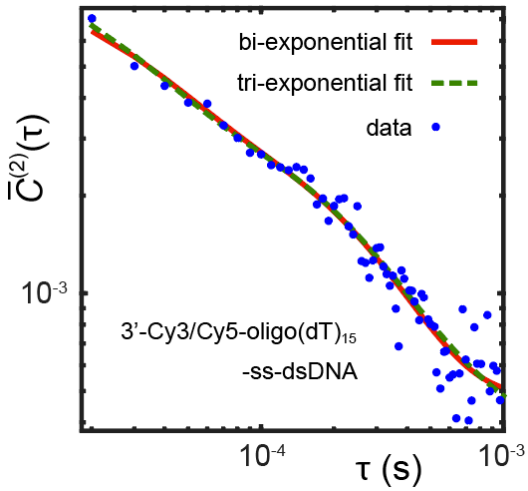

Figure S5. Comparison between model fits of the experimental two-point TCF to bi-exponential and tri-exponential functions. The bi-exponential function describes the decay to an acceptable level of accuracy.

Table S3. Results of numerical regression analysis of fitting multi-exponential model functions to experimental two-point TCFs for the four different ss-dsDNA constructs investigated in this work. The model functions are sums of exponentials: $y = \Sigma_{i=1}^{A} \alpha_i e^{-t/\gamma_i}$, where A = 1, 2 or 3 to find the α_i and γ_i that maximize the R^2 value. An example comparison between the best fits of a bi-exponential (A = 2) and tri-exponential (A = 3) model function to the experimental two-point TCF of the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct is shown in Fig. S5. The values listed are the goodness-of-fit parameter R^2. For each of the ss-dsDNA constructs, the R^2 value does not improve significantly upon increasing the complexity of the model function from a bi-exponential to a tri-exponential decay.

Construct	Mono-exponential	Bi-exponential	Tri-exponential
3'-oligo(dT)$_{15}$	0.939	0.980	0.982
3'-oligo(dT)$_{14}$	0.912	0.963	0.968
5'-oligo(dT)$_{15}$	0.925	0.937	0.944
5'-oligo(dT)$_{14}$	0.900	0.934	0.937

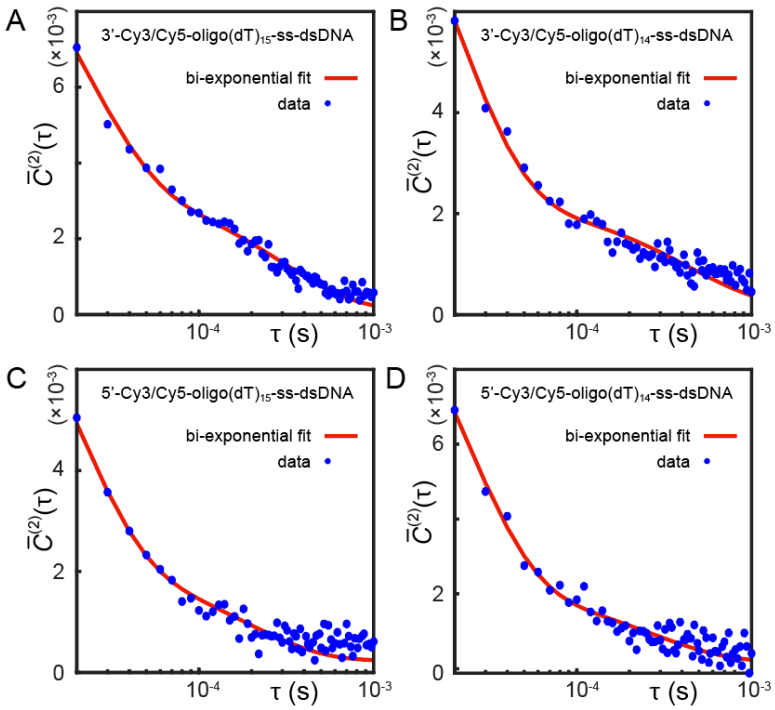

Figure S6. Two-point TCFs of the four ss-dsDNA constructs. (***A***) 3'-Cy3/Cy5-oligo(dT)$_{15}$; (***B***) 3'-Cy3/Cy5-oligo(dT)$_{14}$; (***C***) 5'-Cy3/Cy5-oligo(dT)$_{15}$; and (***D***) 5'-Cy3/Cy5-oligo(dT)$_{14}$. Red curves are model bi-exponential functions $\bar{C}^{(2)}(\tau) = \mathcal{A}_2 e^{\lambda_2 \tau} + \mathcal{A}_3 e^{\lambda_3 \tau} + B$, which are based on the optimized values obtained using the three-state master equation (see Table S4).

Table S4. Optimized parameters for the model bi-exponential functions $\bar{C}^{(2)}(\tau) = \mathcal{A}_2 e^{\lambda_2 \tau} + \mathcal{A}_3 e^{\lambda_3 \tau} + B$ corresponding to the red curves shown in Fig. S6.

Construct	$\mathcal{A}_2$ ($\times 10^{-3}$)	$-\lambda_2^{-1}$ (μs)	$\mathcal{A}_3$($\times 10^{-3}$)	$-\lambda_3^{-1}$ (μs)	B ($\times 10^{-4}$)
3'-oligo(dT)$_{15}$	3.65	273	8.76	20.3	1.24
3'-oligo(dT)$_{14}$	2.34	369	10.0	18.1	1.66
5'-oligo(dT)$_{15}$	2.02	188	8.93	17.6	2.37
5'-oligo(dT)$_{14}$	2.26	230	11.8	20.8	1.92

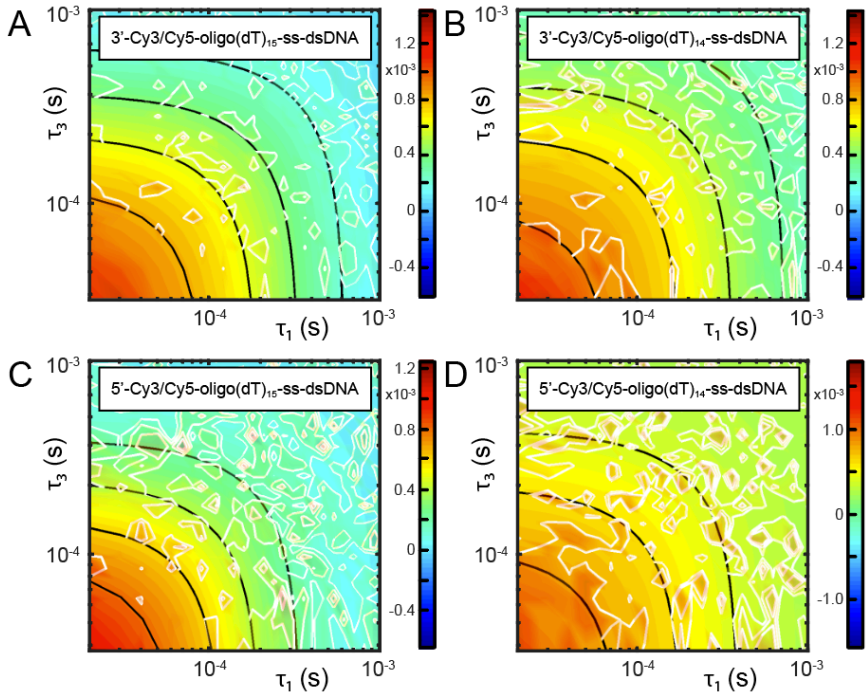

Figure S7. Overlay of experimental and simulated four-point TCFs of the four ss-dsDNA constructs, which are shown as two-dimensional contour plots. (***A***) 3'-Cy3/Cy5-oligo(dT)$_{15}$; (***B***) 3'-Cy3/Cy5-oligo(dT)$_{14}$; (***C***) 5'-Cy3/Cy5-oligo(dT)$_{15}$; and (***D***) 5'-Cy3/Cy5-oligo(dT)$_{14}$. Black curves are model functions $\bar{C}^{(4)}(\tau_1, \tau_2 = 0, \tau_3) = \mathcal{A}_{22}e^{\lambda_2(\tau_1+\tau_3)} + \mathcal{A}_{23}e^{\lambda_2\tau_1+\lambda_3\tau_3} + \mathcal{A}_{32}e^{\lambda_3\tau_1+\lambda_2\tau_3} + \mathcal{A}_{33}e^{\lambda_3(\tau_1+\tau_3)} + D$, which are based on the optimized values obtained using the three-state master equation (see Table S5).

Table S5. Optimized parameters for the model functions $\bar{C}^{(4)}(\tau_1, \tau_2 = 0, \tau_3) = \mathcal{A}_{22}e^{\lambda_2(\tau_1+\tau_3)} + \mathcal{A}_{23}e^{\lambda_2\tau_1+\lambda_3\tau_3} + \mathcal{A}_{32}e^{\lambda_3\tau_1+\lambda_2\tau_3} + \mathcal{A}_{33}e^{\lambda_3(\tau_1+\tau_3)} + D$, corresponding to the black contour lines shown in Fig. S7. Note that the eigenvalues, λ_2 and λ_3, are the same as those listed in Table S4.

Construct	$\mathcal{A}_{22}$ $(\times 10^{-4})$	$\mathcal{A}_{33}(\times 10^{-4})$	$\mathcal{A}_{23} = \mathcal{A}_{32}(\times 10^{-5})$	D $(\times 10^{-5})$
3'-oligo(dT)$_{15}$	8.53	2.82	-2.86	5.56
3'-oligo(dT)$_{14}$	5.65	9.21	-2.15	5.30
5'-oligo(dT)$_{15}$	6.40	5.46	-2.79	1.26
5'-oligo(dT)$_{14}$	4.07	3.98	-2.17	1.35

Table S6. Optimized forward and backward time constants corresponding to the elementary chemical steps obtained for the four ss-dsDNA constructs studied in this work. The time constant parameters are defined in Eq. (9) of the main text. Error bars are on the order of the last significant figure, and are listed for the 3'-oligo(dT)$_{15}$ construct in Table S11.

Construct	k_{12}^{-1} (ms)	k_{21}^{-1}(ms)	k_{13}^{-1} (ms)	k_{31}^{-1} (ms)	k_{23}^{-1} (ms)	k_{32}^{-1} (ms)
3'-oligo(dT)$_{15}$	0.113	0.025	104	1.85	4.01	0.325
3'-oligo(dT)$_{14}$	0.203	0.020	155	1.55	4.80	0.488
5'-oligo(dT)$_{15}$	0.154	0.020	273	1.80	4.15	0.211
5'-oligo(dT)$_{14}$	0.097	0.027	137	2.00	4.92	0.262

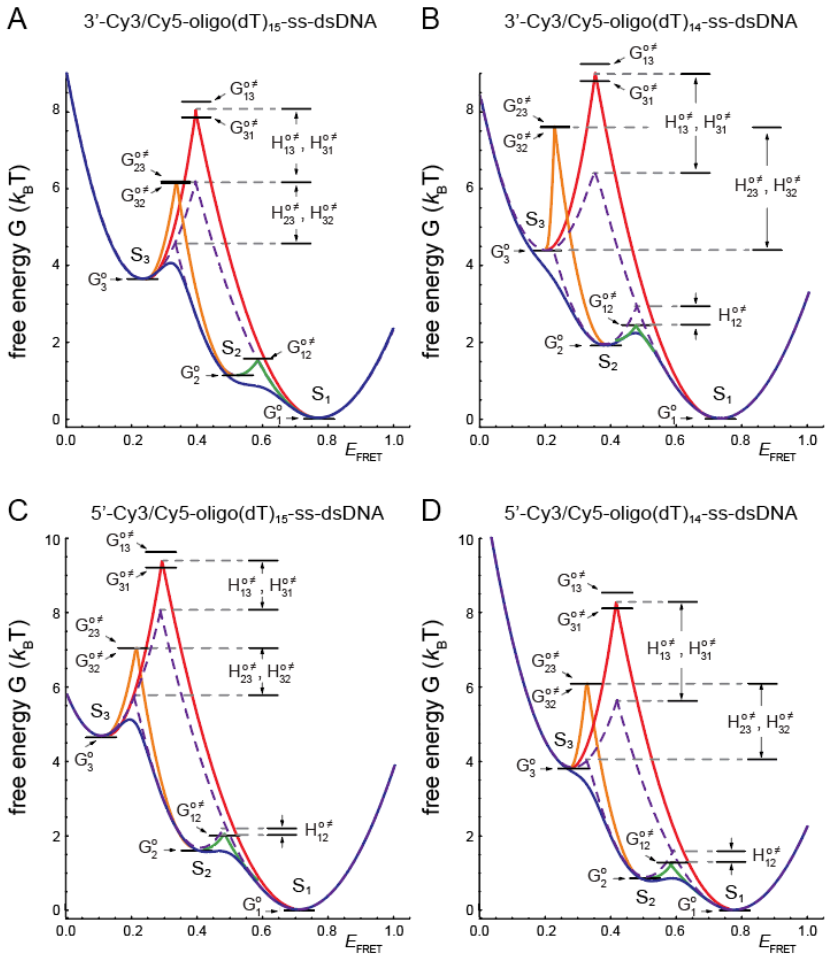

Figure S8. Free energy landscapes constructed from microsecond-resolved smFRET experiments for the four ss-dsDNA constructs: (***A***) 3'-Cy3/Cy5-oligo(dT)$_{15}$; (***B***) 3'-Cy3/Cy5-oligo(dT)$_{14}$; (***C***) 5'-Cy3/Cy5-oligo(dT)$_{15}$; and (***D***) 5'-Cy3/Cy5-oligo(dT)$_{14}$. The free energy landscape is constructed from the PDFs (Fig.

S3 and Table S1) according to the Boltzmann distribution [Eq. (2) of the main text]. Intersecting dashed purple lines indicate the entropic contributions to the free energy barriers that separate macrostates. Intersecting solid curves indicate the total free energy barriers calculated from the optimized rate constants (Table S6) according to the Arrhenius equation [Eq. (3) of the main text]. Values of the free energy minima and transition barriers are given in Table S7. Estimates to the enthalpic and entropic contributions to the transition barriers are given in Table S8.

Table S7. Optimized free energy minima and transition barriers for the four ss-dsDNA constructs studied in this work. The S_1 macrostate is taken to be the ground state and excited state energies are given by $G_i^o = -k_B T ln\left(p_i^{eq}/\sigma_i\sqrt{2\pi}\right) - G_1$ [Eq. (2) of the main text] with parameters p_i^{eq} and σ_i listed in Table S1. The activation energies are determined by $G_{ij}^{o\ddagger} = -ln\left(k_{ij}/k_{max}\right)$ [Eq. (3) of the main text, with $k_{max} = k_{21}$ the fastest time constant of the system]. The rate constant parameters are listed in Table S6. Energies are listed in units of $k_B T$.

Construct	G_1^o	G_2^o	G_3^o	$G_{12}^{o\ddagger}$	$G_{21}^{o\ddagger}$	$G_{13}^{o\ddagger}$	$G_{31}^{o\ddagger}$	$G_{23}^{o\ddagger}$	$G_{32}^{o\ddagger}$
3'-oligo(dT)$_{15}$	0	1.11	3.62	1.52	0	8.34	4.30	5.08	2.56
3'-oligo(dT)$_{14}$	0	1.91	4.20	2.32	0	8.96	4.35	5.48	3.19
5'-oligo(dT)$_{15}$	0	1.64	4.62	2.04	0	9.52	4.50	5.34	2.36
5'-oligo(dT)$_{14}$	0	0.89	3.82	1.28	0	8.53	4.30	5.20	2.27

Table S8. Transition barrier enthalpies and entropies for interconversion between macrostates S_1, S_2 and S_3. Entropies The entropic contribution to the transition barrier is defined $-TS_{ij}^{o\ddagger} = G_{ij}^{\text{intersect}}$ with $G_{ij}^{\text{intersect}}$ the free energy at the point of intersection between the parabolas describing macrostate-*i* and macrostate-*j* (dashed purple curves in Fig. 4*D* of the main text). The enthalpy of transition is estimated to be $H_{ij}^{o\ddagger} = G_{ij}^{o\ddagger} - G_{ij}^{\text{intersect}}$. Enthalpic and entropic contributions to the transition barriers are given in units of $k_B T$.

Construct	$-TS_{12}^{o\ddagger}$	$-TS_{13}^{o\ddagger}$	$-TS_{23}^{o\ddagger}$	$H_{12}^{o\ddagger}$	$H_{21}^{o\ddagger}$	$H_{13}^{o\ddagger}$	$H_{31}^{o\ddagger}$	$H_{23}^{o\ddagger}$	$H_{32}^{o\ddagger}$
3'-oligo(dT)$_{15}$	1.50	6.28	4.65	0.02	0	2.06	1.64	1.54	1.53
3'-oligo(dT)$_{14}$	2.90	6.45	4.25	-0.58	0	2.51	2.10	3.14	3.14
5'-oligo(dT)$_{15}$	2.25	7.95	5.70	-0.21	0	1.57	1.17	1.28	1.28
5'-oligo(dT)$_{14}$	1.55	5.65	4.00	-0.27	0	2.88	2.47	2.09	2.09

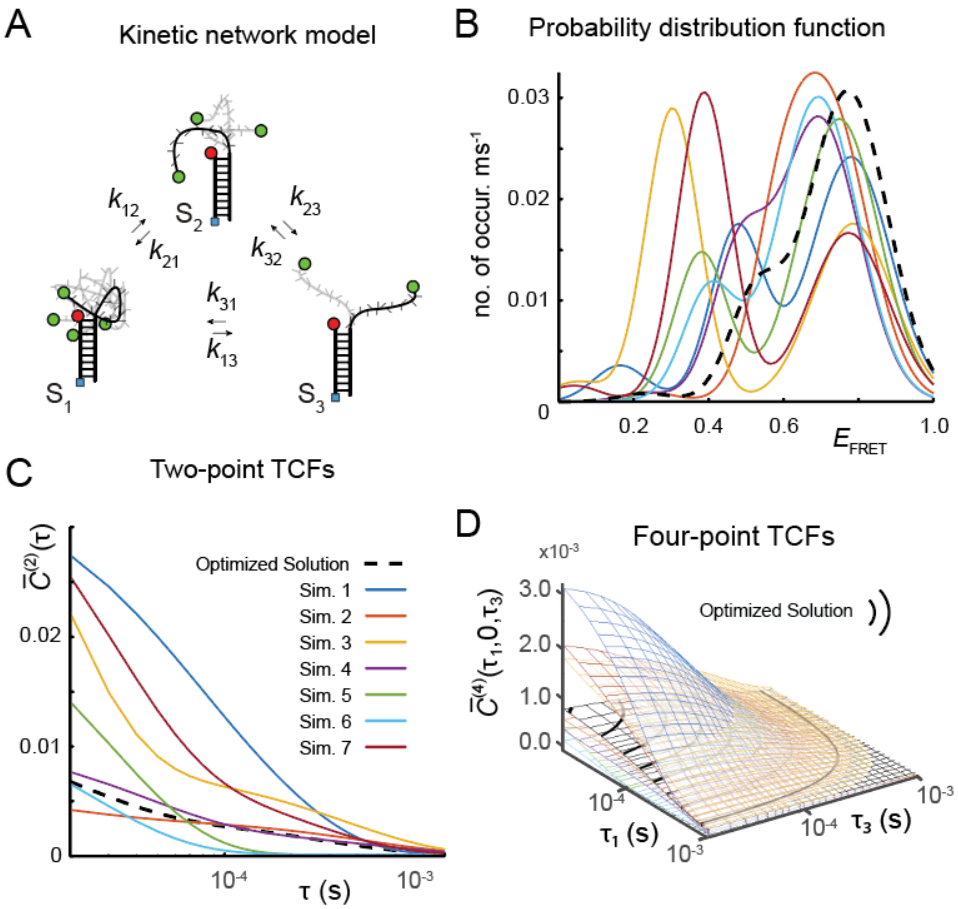

Figure S9. Comparison between the data-optimized solution to the three-state master equation for the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct and various non-optimized solutions (labeled simulations). (***A***) Kinetic scheme depicting the elementary chemical steps connecting the three macrostates of the oligo(dT)$_{15}$ template: 'compact' S_1, 'partially-extended' S_2, and 'highly-extended' S_3. (***B***) Simulated probability distribution functions (PDFs). (***C***) Simulated two-point time correlation functions (TCFs). (***D***) Simulated four-point TCFs. The comparison serves to demonstrate the sensitivity of the analysis to the choice of model input parameters. The values of the time constant parameters and macrostate FRET efficiencies used for the simulations are listed in Table S9.

Table S9. Parameters used for the simulated PDFs and TCFs shown in Fig. S9.

Trial	k_{12}^{-1} (ms)	k_{21}^{-1} (ms)	k_{13}^{-1} (ms)	$k_{31}^{-1}1$ (ms)	k_{23}^{-1} (ms)	k_{32}^{-1} (ms)	$\langle E_{FRET}\rangle_1$	$\langle E_{FRET}\rangle_2$	$\langle E_{FRET}\rangle_3$
Opt. Sol.	0.113	0.025	104	1.85	4.01	0.325	0.772	0.524	0.233
Sim. 1	0.148	0.070	12.5	1.24	0.742	0.155	0.780	0.476	0.164
Sim. 2	0.086	0.021	160	3.41	5.21	0.452	0.716	0.585	0.238
Sim. 3	0.028	0.031	32.8	2.43	5.51	0.372	0.786	0.303	0.047
Sim. 4	0.173	0.056	175	3.40	7.57	0.457	0.695	0.478	0.238
Sim. 5	0.118	0.042	129	2.35	5.86	0.030	0.748	0.380	0.108
Sim. 6	0.125	0.031	155	0.05	4.07	0.051	0.693	0.402	0.140
Sim. 7	0.0.54	0.067	54.8	3.47	4.70	0.244	0.774	0.388	0.040

Numerical Calculations using the Three-State Master Equation. We adopted a kinetic network model that is described by the master equation [Eq. (8) of the main text] with the number of macrostates $M = 3$. We assumed a cyclical kinetic scheme, as shown in Fig. S9*A*. The six forward and backward rate constants obey the detailed balance condition (8, 9)

$$k_{32} = k_{31}\, k_{12}\, k_{23}/k_{21}k_{13} \tag{S4}$$

In order to model the statistical functions constructed from our data (two-point and four-point TCFs and PDFs), it is necessary to determine the six rate constants in addition to the values of each of the macrostate's mean FRET efficiencies, $\langle E_{FRET}\rangle_i$ with $i \in \{1,2,3\}$, which leads to a total of eight independently adjustable parameters in our model.

The rate matrix, $\boldsymbol{K}$ [Eq. (9) of the main text] is written:

$$\boldsymbol{K} = \begin{bmatrix} -(k_{12}+k_{13}) & k_{21} & k_{31} \\ k_{12} & -(k_{21}+k_{23}) & k_{32} \\ k_{13} & k_{23} & -(k_{31}+k_{32}) \end{bmatrix} \tag{S5}$$

The eigenvalues and eigenvectors of $\boldsymbol{K}$ provide the relaxation times and collective modes of the coupled system of elementary chemical reactions. The eigenvalues and eigenvectors of $\boldsymbol{K}$ are used in Eq. (12) to determine the state-to-state conditional probabilities, $p_{ij}(t)$ with $i, j \in \{1,2,3\}$, and the equilibrium probability, p_i^{eq}, for each macrostate. The equilibrium probabilities, together with the mean FRET efficiencies, were used to determine the PDF according to: $p(E_{FRET}) = \sum_{j=1}^{M} p_j(E_{FRET})$ [see discussion around Eq. (2)]. The conditional probabilities and the mean FRET efficiencies were used in Eq. (5) and Eq. (7) to calculate the two-point and four-point TCFs, respectively. In Fig. S9*B*, we show a set of simulated PDFs for randomly chosen sets of rate constants and mean FRET efficiencies (listed in Table S9). Simulated two-point and four-point TCFs corresponding to the same sets of parameters are shown in Fig. S9*C* and Fig. S9*D*, respectively. The comparison between the simulated functions and those corresponding to the optimized solutions (shown as solid black dashed curves in Fig. S9) serves to demonstrate the sensitivity of this analysis to the choice of model input parameters.

Solving the Master Equation using the Similarity Transform Method. We applied a computationally efficient means to obtain solutions to the master equation by using the similarity transformation discussed in Sec. 3.4 (10):

$$p_n(t) = U\left[e^{\lambda t}\right]U^{-1}p_n(0) \tag{S6}$$

In Eq. (S6), the unitary matrix $\boldsymbol{U} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_M]$ has columns that are the eigenvectors of the rate matrix $\boldsymbol{K}$. The matrix $\left[\boldsymbol{e}^{\boldsymbol{\lambda} t}\right]$ is diagonal with non-zero elements $\left[e^{\lambda t}\right]_{ii} = e^{\lambda_i t}$, where the decay constants λ_1, λ_2 and λ_3 are the eigenvalues of $\boldsymbol{K}$.

After imposing the condition $\lambda_1 = 0$ (discussed in Sec. 3.4 of the main text), the matrix $\left[\boldsymbol{e}^{\boldsymbol{\lambda} t}\right]$ can be written:

$$\left[\boldsymbol{e}^{\boldsymbol{\lambda} t}\right] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{\lambda_2 t} & 0 \\ 0 & 0 & e^{\lambda_3 t} \end{bmatrix}$$

We rewrite Eq. (S6) in terms of the conditional probability matrix $\boldsymbol{p}(t) = [\boldsymbol{p}_1, \boldsymbol{p}_2, \dots, \boldsymbol{p}_M]$:

$$\boldsymbol{p}(t) = \boldsymbol{U}\left[\boldsymbol{e}^{\boldsymbol{\lambda} t}\right]\boldsymbol{U}^{-1}\boldsymbol{p}(0) \tag{S7}$$

Upon setting the initial condition matrix $\boldsymbol{p}(0)$ equal to the identity matrix $\boldsymbol{I}$, which represents all possible initial conditions, the matrix elements of $\boldsymbol{p}(t)$ become the conditional probabilities $p_{ij}(t)$ according to:

$$\begin{bmatrix} p_{11} & p_{21} & p_{31} \\ p_{12} & p_{22} & p_{32} \\ p_{13} & p_{23} & p_{33} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{\lambda_2 t} & 0 \\ 0 & 0 & e^{\lambda_3 t} \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{S8}$$

Iterative Multidimensional Optimization Procedure. We performed an iterative multi-parameter optimization procedure, similar to the one described by Phelps *et al*. (6, 9), to find the set of kinetic and thermodynamic parameters that best matches the experimentally-derived PDFs and two-point and four-point TCFs.

We introduce as a metric the weighted error function, χ^2, to assess the quality of agreement between the master equation solutions and the statistical functions we have constructed from our experimental data. We define the error function as the cumulative squared-difference between the measured and simulated quantities, represented as y and g, respectively:

$$\chi^2 = \sum_{q=1}^{3} w_q(\tau, \tau_1, \tau_2, \tau_3) \sum_i [y_i - g_i(\{k_{lm}, \langle E_{FRET} \rangle_l\})]^2 \tag{S9}$$

In Eq. (S9), q is an index that enumerates: (1) the FRET histogram; (2) the two-point TCF; and (3) the four-point TCF, and i is the domain index for each function. We chose the weighting coefficients, w_n, so that the histogram and the TCFs contributed approximately equally to the error function. In addition, w_n is a function of the lag times $\tau, \tau_1, \tau_2, \tau_3$, such that the TCFs are assigned more weight at earlier times, as we did in our previous studies (6).

Figure S10. Workflow diagram of the multi-parameter optimization procedure. In order to effectively search the parameter space to obtain an optimized solution to the master equation, we applied a numerical 'genetic algorithm' (GA) with principles analogous to evolution by natural selection. The GA consists of four sequential steps, which are applied iteratively to successive generations of a 'population' of distinct solutions to the master equation. (*i*) In the 'initialization step,' the first-generation population is generated as a random set of trial solutions with mixed levels of fitness, as determined by the χ^2 target function [see discussion of Eq. (S9)]. A population of solutions is represented in the far-left column with different colors indicating varying levels of fitness. Each 'individual' member of the population represents a numerical array containing a specific set of model parameters (i.e., FRET efficiency values and time constant parameters), which corresponds to a unique solution to the master equation; (*ii*) In the 'selection step,' the individuals are ranked according to their fitness. The fittest individuals (typically chosen as the top 50%, shaded green) are retained without alteration for the next stage of the GA, while the least fit individuals (bottom 50%, shaded red) are discarded from the population. The highest tiers of fitness are selected (typically 25%, shaded blue) for 'reproduction,' which involves the random exchange (or 'mixing') of parameters between individuals. The transformed population is represented in the third column with the reproductive fraction represented by the light-blue shaded region; (*iii*) In the 'mutation step,' the parameters of all but the highest-fitness individual (shaded orange) are randomly mutated to varying degrees as a means to ensure continued 'genetic diversity.' The final population is indicated in the far-right column with mutated individuals shaded purple; (*iv*) In the final 'evaluation step,' the fitness of each individual is calculated, and the one with the highest fitness is selected as the new optimized solution. If the fitness of the optimized solution is improved relative to that of the previous generation, the new generation is used as input to repeat the procedure. If the fitness of the optimized solution exceeds a preset threshold value specified by the parameter η, the loop is terminated. The output of the GA is then further refined using a multi-parameter linear least squares regression algorithm (MATLAB *patternsearch*) to obtain the final optimized solution.

The output parameters of the model are: (*i*) the set of time constants, k_{ij}^{-1} with $j \in \{1,2,3\}$, which describe the transfer times between macrostate-i and macrostate-j and (*ii*) the mean FRET efficiency values corresponding to the three macrostates, $\langle E_{FRET} \rangle_i$ with $i \in \{1,2,3\}$. We implemented a genetic algorithm (GA) to broadly search the parameter space and to obtain an approximate set of parameters that produced a relatively low value of χ^2 using an arbitrary set threshold value η. A schematic workflow diagram for the GA optimization procedure is shown in Fig. S10. The output solution obtained from the GA was then fed into the nonlinear multivariable problem solver, *patternsearch* (MATLAB, The MathWorks, USA) to arrive at a final optimized solution. The entire procedure was repeated iteratively until the final solution was obtained. The final optimized solutions are those presented in Table S1 and Table S6.

Statistical Uncertainty Analysis. We performed an analysis of the statistical uncertainty for each of the free parameters used in our model for the case of the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct (see Fig. S11 and Table S10). This analysis is based on calculating the relative deviation of the error function, $\Delta\chi^2/\chi_0^2$, from the optimized value, χ_0^2, as a function of the time constant and FRET efficiency parameter uncertainties. The error function cross-sections given by Eq. (S9), and shown in Fig. S11, demonstrate that the optimized values we have obtained for the free parameters each correspond to a stable minimum. We assign the statistical uncertainty of these values to a 1% deviation of the error function relative to the corresponding optimized value. From this analysis, we obtain error bars for the free parameters, which we list in Table S10.

Table S10. Error bars determined from the cross-sections of the error function surfaces $\Delta\chi^2/\chi_0^2$, shown in Fig. S11.

Δk_{12}^{-1} (ms)	Δk_{21}^{-1} (ms)	Δk_{13}^{-1} (ms)	$\Delta k_{31}^{-1}1$ (ms)	Δk_{23}^{-1} (ms)	$\Delta\langle E_{FRET}\rangle_1$	$\Delta\langle E_{FRET}\rangle_2$	$\Delta\langle E_{FRET}\rangle_3$
±0.004	±0.0005	±1.85	±0.03	±0.03	±0.003	±0.004	±0.005

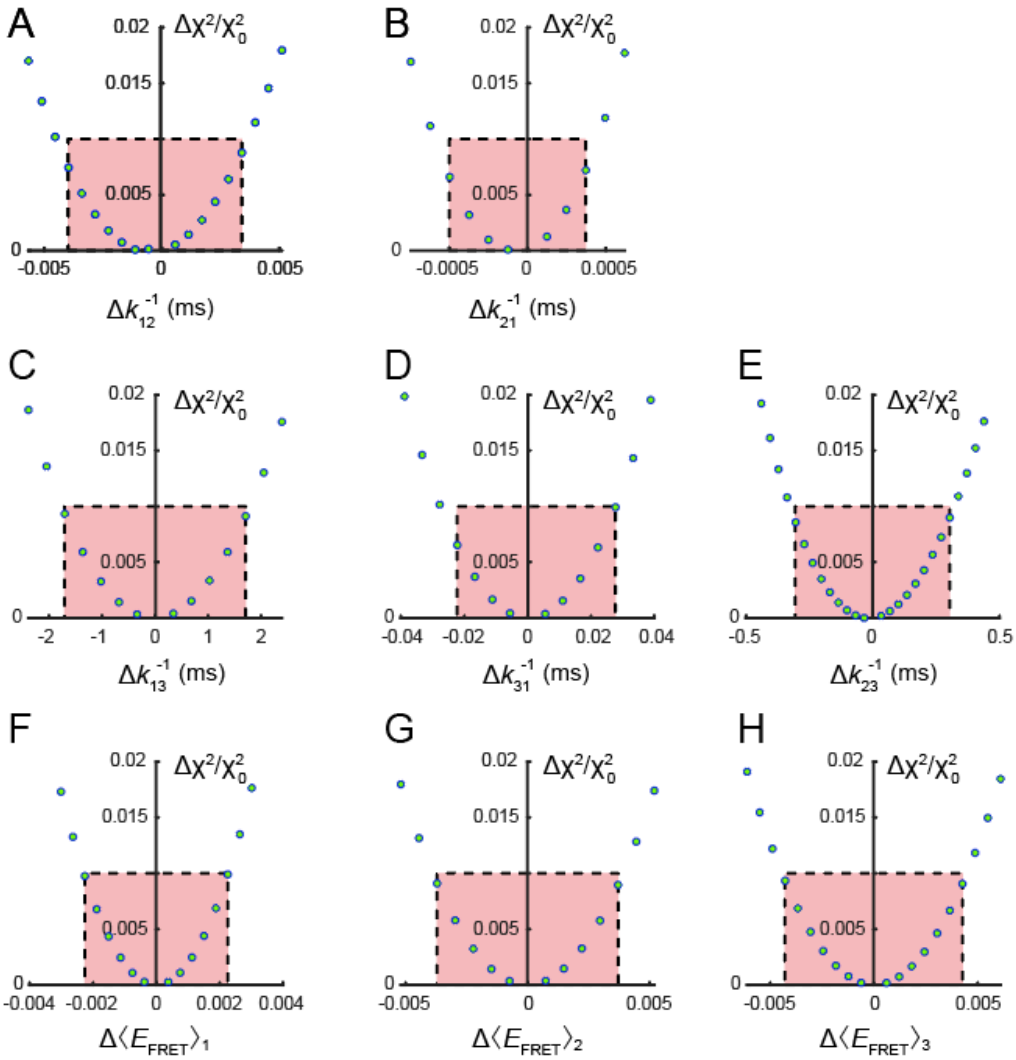

Figure S11. Relative deviation of the error function, $\Delta\chi^2/\chi_0^2$, from the optimized value, χ_0^2, as a function of the time constant and FRET efficiency parameter uncertainties. Cross-sections of the error function given by Eq. (S9) are shown for the optimization to the 3'-Cy3/Cy5-oligo(dT)$_{15}$-ss-dsDNA construct data set, and for the uncertainties (***A***) Δk_{12}^{-1}, (***B***) Δk_{21}^{-1}, (***C***) Δk_{13}^{-1}, (***D***) Δk_{31}^{-1}, (***E***) Δk_{23}^{-1}, (***F***) $\Delta\langle E_{\mathrm{FRET}}\rangle_1$, (***G***) $\Delta\langle E_{\mathrm{FRET}}\rangle_2$, and (***H***) $\Delta\langle E_{\mathrm{FRET}}\rangle_3$. Here $\Delta x = x - x_0$, and x_0 is the value corresponding to the optimized set of parameters, which corresponds to the minimum of the multidimensional parameter surface. The red-shaded rectangles indicate the 'confidence region' based on a 1% relative deviation, which we associate with the quality of the experimental data. A similar analysis was performed for the data obtained for other ss-dsDNA constructs, and the cross-sections are similar to those shown above. The associated error bars reported in Table S10 are rounded up to two significant figures. We note that the time constants associated with Panels (*A*) – (*E*) are related by the detailed balance condition [Eq. (S4)], and the error bars of forward and backward elementary chemical steps were determined accordingly.

References

1. S. D. Chandradoss, A. C. Haagsma, Y. K. Lee, J.-H. Hwang, J.-M. Nam, and C. Joo, Surface passivation for single-molecule protein studies. *J. Vis. Exper.* **86**, e50549 (2014).
2. A. T. Martilla, O. H. Laitinen, K. J. Airenne, T. Kulik, E. A. Bayer, M. Wilchek, and M. S. Kulomaa, Recombinant neutralite avidin: a non-glycosylated, acidic mutant of chicken avidin that exhibits high affinity for biotin and low non-specific binding properties. *FEBS Lett.* **467**, 31-36 (2000).
3. C. Joo, S. A. McKinney, M. Nakamura, I. Rasnick, S. Myong, and T. Ha, Real-time observation of RecA filament dynamics with single monomer resolution. *Cell* **126**, 515-527 (2006).
4. I. Rasnik, S. A. McKinney, and T. H., Nonblinking and long-lasting single-molecule fluorescence imaging. *Nature Meths.* **3**, 891-893 (2006).
5. C. Phelps, W. Lee, D. Jose, P. H. von Hippel, and A. H. Marcus, Single-molecule FRET and linear dichroism studies of DNA 'breathing' and helicase binding at replication fork junctions. *Proc Natl Acad Sci U S A* **110**, 17320 – 17325 (2013).
6. C. Phelps, B. Israels, D. Jose, M. C. Marsh, P. H. von Hippel, and A. H. Marcus, Using microsecond single-molecule FRET to determine the assembly pathways of T4 ssDNA binding protein onto model DNA replication forks. *Proc. Nat. Acad. Sci. Plus USA* **114**, E3612-E3621 (2017).
7. D. V. O'Connor, and D. Phillips, *Time-Correlated Single Photon Counting* (Academic Press, 1984).
8. L. E. Reichl, *A Modern Course in Statistical Physics* (John Wiley & Sons, New York, 1998).
9. C. Phelps, B. Israels, D. Jose, M. C. Marsh, P. H. von Hippel, and A. H. Marcus, Using multi-order time correlation functions (TCFs) to elucidate biomolecular reaction pathways from microsecond single-molecule fluorescence experiments. *J. Phys. Chem. B* **120**, 13003-13016 (2016).
10. G. Beddard, *Applying Maths in the Chemical and Biological Sciences: An example-based approach* (Oxford University Press, Oxford, 2009).