Connecting Rare DNA Conformations and Surface Dynamics using Single-Molecule Resonance Energy Transfer

Mark Kastantin and Daniel K. Schwartz*

Department of Chemical and Biological Engineering

University of Colorado Boulder

Boulder, Colorado 80309

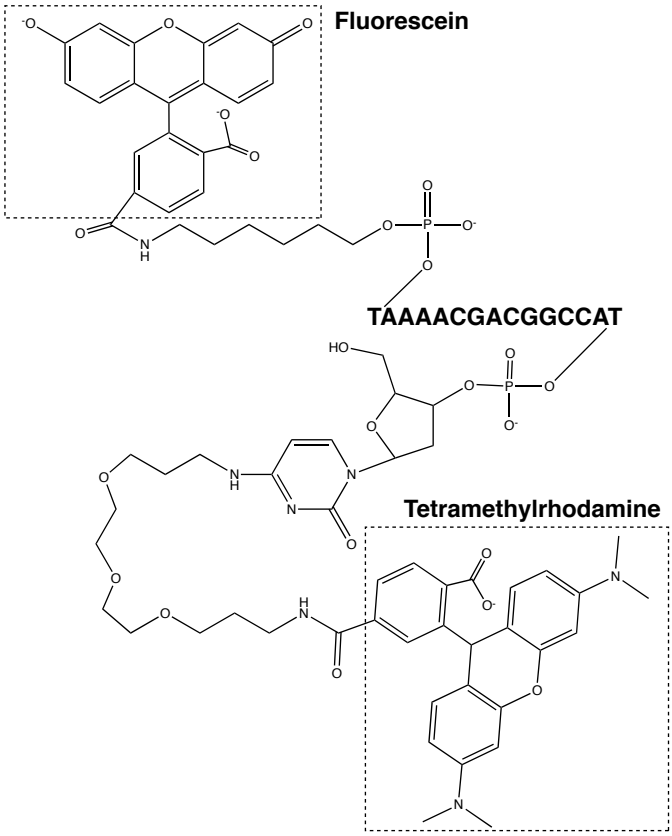
*To whom correspondence should be addressed

daniel.schwartz@colorado.edu

Supporting Information

Structure of dual-labeled ssDNA

The structure of end-labeled ssDNA is shown in Supporting Figure 1. The ssDNA sequence is given by its letter code while the structure of dyes and the linkers that connect them to ssDNA ends are drawn explicitly. Note that the contour length of the 15 base pair ssDNA is much longer than that of the linkers.

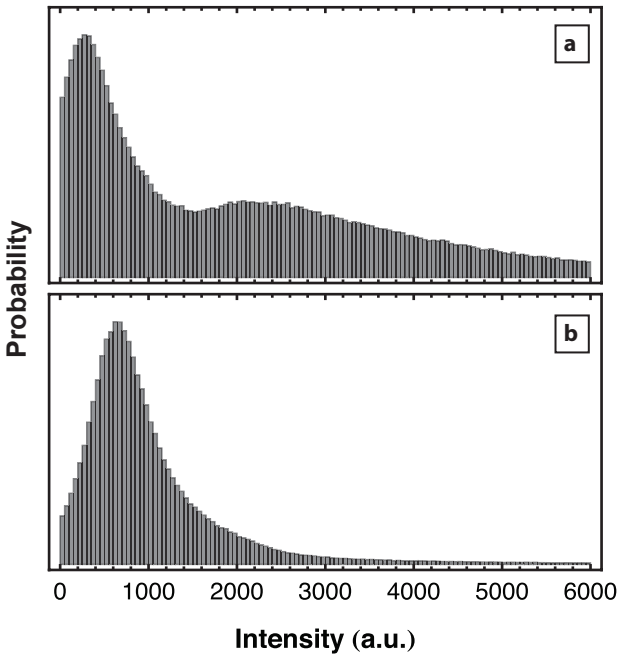


Supporting Figure S1. Structure of end-labeled ssDNA.

Determination of direct excitation of TMR

The intensity histogram for all steps of all trajectories observed in these experiments is shown in Supportion Figure 2. In the acceptor channel, the large peak near 0 intensity represents direct excitation of tetramethylrhodamine (TMR) from the 491 nm laser whereas the broad peak at higher intensity represents a distribution of energy transfer events. In the donor channel, only

one peak is observed. This is the expected behavior for the donor species where energy transfer away from FAM serves to shift the direct excitation peak to lower intensities but not to create a second peak.



Supportingary Figure S2. Intensity distribution for all steps of all trajectories. (a) Acceptor channel (b) Donor channel

Surface residence time distribution of ssDNA

Surface residence times were calculated as the number of frames on which the object was identified, multiplied by the exposure time of each frame. The uncertainty in assigning the residence time of an object is assumed to be the acquisition time divided by $\sqrt{2}$. This is due to the fact that an object may still be identified during an acquisition in which it is present for a large fraction of the time and conversely, the object can be ignored if it is only present for a small fraction of the acquisition time.

The experimental cumulative residence time distribution was constructed by first counting the number of objects observed to have a surface residence time greater than a given time, with error that was expected to follow Poisson statistics. These data were then corrected for the finite

length of a movie whereby the adsorption and desorption of an object has a lower *a priori* probability of being observed for longer residence times because there are fewer opportunities to observe both events in a finite window. Further details concerning this correction have been provided elsewhere.[1]

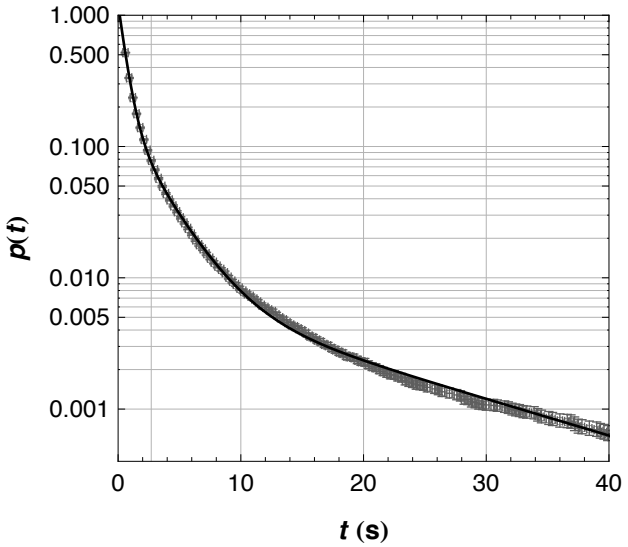
The surface residence time of a given ssDNA population is assumed to follow first-order desorption kinetics and consequently the integrated or cumulative residence time distribution is described as the sum of all such populations as shown in equation (S1):

$$\text{(S1)} \qquad p(t) = \frac{\int_t^\infty \sum_i f_i e^{-\frac{t'}{\tau_i}} \Big/ \tau_i \, dt'}{\int_0^\infty \sum_i f_i e^{-\frac{t'}{\tau_i}} \Big/ \tau_i \, dt'} = \sum_i f_i e^{-\frac{t}{\tau_i}}$$

Here, $p(t)$ is the probability that a given object will have a residence time greater than or equal to time t. Each population is denoted with the subscript, i, and τ_i is the inverse of the first-order desorption rate constant for that population (*i.e.* that population's mean surface residence time). The relative fraction of all analyzed objects represented by population i is f_i where $\Sigma f_i = 1$.

The cumulative residence time distribution for ssDNA in this work is shown in Figure S3 along with a tri-exponential fit using equation S1. Three populations (A-C) were determined using the same criteria for determining the number of diffusive modes in equation (3) as discussed in the Methods section. The relative fractions of the populations, along with their characteristic residence times are given in Table S1. The short-lived population, with $\tau = 0.57$s, comprises 86% of the objects observed. The population with the next shortest characteristic residence time has $\tau = 2.6$s. Thus, the choice of 2.7s is made for a minimum residence time of objects included in this work in order to exclude objects from the short-lived population. The majority of objects in

this short-lived population will have post-adsorption behavior that couples to their pre-desorption behavior, complicating analyses of each process.



Supporting Figure S3. The cumulative residence time distribution is shown for ssDNA on APTES functionalized fused silica. The data is shown in grey with error bars in both directions. The tri-exponential fit to this distribution is shown with the parameters given in Table S1.

Table S1. Parameters used to fit equation S1 to the cumulative residence time distribution.

Population	f_i	τ_i
A	0.856(10)	0.57(2)
B	0.137(6)	2.6(1)
C	0.0007(2)	16(1)

End-to-end distances of self-avoiding walks

This work summarizes the work of Bishop and Clarke, which gives the probability distribution of end-to-end distances for self-avoiding walks (SAWs) in two and three dimensions.[2] Their work performs Monte Carlo simulations of SAWs and compares the resulting end-to-end distance distribution to several different analytical forms. A scaling form, developed by des Cloizeaux,[3] was found to best represent the simulated data and is given below:

$$\text{(S2)} \qquad P_\delta(\tilde{r}) \approx \tilde{r}^{\theta} \exp\left(-\tilde{r}^{t}\right)$$

In equation S2, θ and *t* are dimension-specific constants and in two dimensions, θ = 4/9 and *t* = 4.0, while in three dimensions, θ = 5/18 and *t* = 2.5. P_δ is the probability of observing $\tilde{r} = r / R_g$ where $R_g = b\, N^{3/(\delta+2)}$ is the radius of gyration, *b* = 0.59 nm is the monomer repeat unit, *N* = 15 is the number of monomers, and δ is the spatial dimension. For δ = 2 or 3, the radius of gyration is 4.5 or 3.0 nm, respectively. The fraction of configurations with an end-to-end distance greater than $\tilde{r}$ is given by equation S3:

$$\text{(S3)} \qquad \begin{aligned} x(\tilde{r}) &= \int_{\tilde{r}}^{\infty} P_\delta(\tilde{r}')\, d\tilde{r}' \Big/ \int_0^{\infty} P_\delta(\tilde{r}')\, d\tilde{r}' \\ &= \Gamma\left(\frac{1+\theta}{t}, \tilde{r}^{t}\right) \Big/ \Gamma\left(\frac{1+\theta}{t}, 0\right) \end{aligned}$$

In equation S3, Γ(*y*,*z*) is the incomplete gamma function that satisfies $\Gamma(y,z) = \int_z^{\infty} s^{y-1} e^{-s}\, ds$. Assuming that the lower limit of detection in our system is 1.0 nm, this gives a value of $\tilde{r}$ = 0.22 in two dimensions and $\tilde{r}$ = 0.33 in three dimensions. From equation S3, with dimension-specific values of θ and *t*, we find that *x* = 0.87 in two dimensions and *x* = 0.73 in three dimensions. Thus the majority of SAW conformations of adsorbed ssDNA, whose spatial dimension may lie somewhere between two and three dimensions, have end-to-end distances above our lower limit on spatial resolution.

References

1. Kastantin, M.; Langdon, B. B.; Chang, D. L.; Schwartz, D. K. Single-Molecule Resolution of Interfacial Fibrinogen Behavior: Effects of Oligomer Populations and Surface Chemistry. *J. Am. Chem. Soc.* **2011,** *133*, 4975-4983.
2. Bishop, M.; Clarke, J. H. R. Investigation of the End-to-End Distance Distribution Function for Random and Self-Avoiding Walks in 2 and 3 Dimensions. *J. Chem. Phys.* **1991,** *94*, 3936-3942.
3. des Cloizeaux, J. Lagrangian Theory for a Self-Avoiding Random Chain. *Phys. Rev. A* **1974,** *10*, 1665-1669.