Single-molecule FRET TACKLE reveals highly dynamic mismatched DNA-MutS complexes

Lauryn E. Sass, Cherie Lanyi, Keith Weninger, and Dorothy A Erie

Supplemental Information

Supplemental Data and Analysis Methods

Single-molecule FRET traces of mismatched DNA-MutS complexes

Sample single molecule FRET traces for GT-mismatched DNA in the presence of MutS are shown in Figure S1. Results reveal diverse dynamics, with molecules undergoing a number of conformational transitions during the FRET time traces. Some molecules undergo only one or two transitions during the duration of trace (typically limited to ~ 60 seconds by photobleaching of the dyes) (Figure S1 A), revealing slow rates of exchange among different conformational states. Similar dynamics were also observed for these complexes bound to quartz surfaces passivated by coating with PEG (*1*) (Figure S1 B), revealing that these dynamics are not the result of interactions of the DNA-protein complexes with the surface.

Control experiments were performed using GT mismatched DNA singly labeled with either the FRET donor or the FRET acceptor in the presence of MutS. In these experiments, no fluctuations were observed in the intensities of either of the dyes, confirming that anti-correlated signal changes for the dual-labeled GT mismatched DNA is a direct function of energy transfer between the FRET donor and the FRET acceptor due to DNA bending induced by MutS. Traces from these control experiments are shown in Figure S6.

Transition edge detection via Gaussian derivative kernel

A derivative of a Gaussian kernel may serve as an optimal edge detector to distinguish between noise in a data signal and transition events (*2*). Briefly, the Gaussian derivative function in one dimension (time, *t*) is defined as:

$$\frac{dG(t;\sigma)}{dt} = -\frac{\exp\left(\frac{-t^2}{2\sigma^2}\right)\cdot t}{\sqrt{2\pi}\cdot\sigma^3}$$

Transition edges are detected by filtering the signal with a discrete Gaussian derivative kernel (denoted FRET') and searching for extrema in the resulting signal (Figure S2). The filtering operation is performed by multiplying the Fourier transforms of the signal and the filter kernel (*2*). For FRET efficiency as a function of time (t):

$$FRET'(t) = F^{-1}\left[F(FRET(t))\cdot F\left(\frac{dG(t;\sigma)}{dt}\right)\right],$$

where F and F^{-1} represent the Fourier and inverse Fourier transforms, respectively. Maxima in the filtered signal correspond to positive (increasing) edges, while minima correspond to negative (decreasing) edges (Figure S2 c).

The derivative of Gaussian kernel contains a scale parameter, σ, which dictates both the precision of edge detection as well as the degree of noise suppression. At a small scale parameter, the signal is not smoothed, and detection of transition edges is more sensitive. At a large scale parameter, the signal is smoothed, and the location of edges may become less precise. To overcome this trade-off, the signal is convolved using a multi-scale analysis approach. The signal is processed first at a large scale to smooth signal features then re-processed at smaller scales to refine the detection accuracy

of transition edges. This analysis software is available for download at https://www.cs.unc.edu/Research/nano/cismm/download/edgedetector/index.html.

Lifetime examination of states comprising the FRET efficiency distribution

To begin lifetime analysis on a FRET efficiency distribution of states, a 2D FRET efficiency distribution of all the FRET conformations observed is generated (Figure 3F). The FRET efficiency distribution of free GT-mismatched DNA (Figure 3F, black cityscape) reveals a single Gaussian peak centered at FRET 0.24. The breadth of this distribution (2σ) was calculated to be 0.15 to 0.16 from a Gaussian fit and is indicative of the reliability in which a single FRET state can be isolated. The FRET distribution for GT-mismatched DNA in the presence of MutS (Figure 3F, gray bars), in contrast, appears bimodal. Fitting it to a double Gaussian distribution results in the breadth of each Gaussian peak being larger than the breadth expected from noise alone (0.16). This observation suggests that more than two species comprise the 'bimodal' Gaussian distribution.

FRET values across the entire distribution were grouped in small segments of FRET efficiencies (FRET +/- 0.005) and a dwell time distribution was generated for the state (plot of frequency vs. dwell time, Figure 3). If the dwell time distributions for adjacent FRET states converged to the same lifetime or lifetimes (within the expected error determined by the exponential fit of the distribution), those FRET efficiencies were combined as a single state and refit to obtain a combined lifetime or lifetime for the state. This process was repeated for the entire FRET distribution.

The combined dwell times of all FRET efficiencies between 0 and 0.29 ('low FRET' state, less DNA bending) converged to a single exponential with an average τ = 3.9 seconds (Figure 3A). The combined dwell times of all FRET efficiencies from 0.61 to 1.0 ('high FRET' state, more DNA bending) converged to a single exponential with an average τ = 13 seconds (Figure 3C). Although lifetime analysis isolated a single state in this range, transition density analysis reveals transitions among multiple conformations within this range of FRET efficiencies (Figure 4A). Based on these transitions, two states were assigned by FRET efficiency: conformation ***B*** (FRET 0.61 to 0.70) and conformation ***SB*** (FRET 0.71 to 1.0). Dwell time distributions were regenerated for each of these separated states (Figure S3). As expected, the dwell time distributions converged to similar lifetimes, where average τ = 13 seconds for conformation ***B*** and τ = 17 seconds for conformation ***SB*** (Figure S3, Table 1).

Dwell times associated with FRET values between 0.30 and 0.40 converged to a single exponential with an average τ = 3.5 seconds (Figure 3B). Dwell times associated with FRET values from 0.41 to 0.50 were best fit to a double exponential decay (n_1 = 279, F_1 = 2771; n_2 = 59, F_2 = 54 (subscripts denoting independent experiments)) and thus comprise two species with average lifetimes τ_1 = 0.53 seconds and τ_2 = 8.7 seconds (Figure 3D). While these species are indistinguishable by FRET efficiency, unique lifetimes indicate that there are two unique states within a single FRET interval (*3*). Finally, a double exponential fit of dwell times associated with FRET values from 0.51 to 0.60 (n_1 = 357, F_1 = 1456; n_2 = 54, F_2 = 3.9) yields two distinct species with average decay constants of τ_1 = 2.2 seconds and τ_2 = 15 seconds (Figure 3E). The state with the shorter lifetime (τ_1 = 2.2 seconds) appears to be unique. However, the similarity of the

longer lifetime species (τ_2) to that of conformation ***B*** in the FRET range 0.61 to 0.70 (15 seconds vs. 13 seconds) suggests that this state is not unique but contains species from conformation ***B***.

Isolating unique transitions from transition density plots

Identifying transitions by FRET efficiency. To identify all transitions between conformational states with different FRET efficiencies, the TDP for all GT-MutS transitions is sliced into FRET ranges corresponding to unique states (Figure 4B, Figure S4). For example, all transitions from conformation ***U*** (section comprised of starting FRET values from 0.29 to 0.41) are extracted from the TDP (Figure 4B). The slice is cross-sectioned to isolate individual transition density peaks resulting from transitions from conformation ***U*** to each other FRET state (Figure 4C), and each 3-dimensional peak is fit to a 2-D Gaussian function (Methods) (*4, 5*). By applying this analysis across the entire transition density distribution, each 2-D Gaussian peak, which corresponds to a given transition, is identified (Figures 2B and S4), and the number of transitions between each of the states is determined.

Results from this analysis directly show that transitions between conformations ***B*** and ***U*** occur more frequently than any other conformational transition and that the most frequent binding transition is ***free DNA*** to ***B***. This analysis also reveals which transitions are least likely to occur, such as those between ***U*** and ***U*/I***. The frequency or infrequency of occurrence of specific transitions, identified by analysis of the TDP, offers clues to unraveling the roles of each of these states in the mode of mismatch recognition by MutS.

As noted in the kinetic lifetime analysis, the transition densities reveal areas reflecting transitions to an additional conformational state at FRET 0.75 (conformation ***SB***, arrows in Figure 4A). Transitions to and from this unique state are clearly identified and shown in Figure S4. This observation confirms that two unique conformational states, ***B*** and ***SB***, comprise FRET efficiencies from 0.61 to 1.0, although these states have nearly indistinguishable lifetimes (Table 1, Figure S3).

Identifying transitions among different species in the same FRET range using lifetimes. The above transition density analysis allows direct observations of transitions between conformational states with *different* FRET efficiencies. However, we were interested in extracting transitions from unique species comprising the same FRET efficiency range but identified by unique lifetimes ('mixed FRET states'). Specifically, what are the transition characteristics of ***U**** vs. ***I*** in the FRET range 0.41 to 0.50 and ***B**** vs. ***B*** in the FRET range 0.51 to 0.60 (Figure 3)?

To separate these distinct species, the double exponential lifetime decay curve is deconvoluted into two single exponential decay curves based on the fit parameters determined from results in Figure 3 (Figures 5A and S5). From this deconvolution, species comprising the short lifetime states are separated from the species comprising the long lifetime states within a given overlapping FRET efficiency state (***U**** from ***I*** in the FRET range 0.41 to 0.50 and ***B**** from ***B*** in the FRET range 0.51 to 0.60) (Figure 5A and S5). Transition density plots (shown as contour images in Figures 5B and S5) are generated for each of these states based on their dwell times, allowing direct observation of transitions to and from each of these unique species. Integration of the two exponential decay curves dictates the amount of misassignment among these states upon

this deconvolution. This integration reveals that conformations ***I*** and ***B*** are contaminated with less than 5% of ***U**** and ***B****, respectively, while conformations ***U**** and ***B**** contain a higher degree of contamination from the longer lifetime states (25% for ***I*** into ***U**** and 45% for ***B*** into ***B****).

This analysis shows that although these species lie within a single FRET efficiency state, their transition characteristics are very different. For example, conformation ***U**** preferentially transitions to conformations ***B*** and ***SB*** (Figure 5C a-e), while conformation ***I*** preferentially results in dissociation of MutS from the DNA (Figure 5C f-j). The transition densities of conformation ***B*** in the FRET range 0.51 to 0.60 (Figure S5) and in the FRET range 0.61 to 0.70 (Figure S4) are nearly identical, confirming that conformation ***B*** occupies FRET efficiencies ranging from 0.51 to 0.70. This result is not surprising because ***B*** is the most populated state. The transition densities shown in Figures 5 and S5 support the claim that each lifetime within the mixed FRET efficiency states represents a unique species with characteristic transition patterns. One limitation to this analysis is that transitions between states ***U**** and ***I*** cannot be identified since they reside at the same FRET efficiency. Similarly, ***B*** and ***B**** have partially overlapping FRET efficiencies, which allows only a fraction of transitions between states to be observed. However, these results show the power of FRET TACKLE in the direct separation of transitions from these different species residing in the same FRET efficiency range.

Relative free energies of DNA-MutS conformational states and transitions

A partition coefficient of individual transitions is calculated from the kinetic rate constants where $Q = k_f / k_r$, where k_f is the forward rate of the transition (for example, ***U*→*B***), and k_r is the reverse rate of the transition (ie. ***B*→*U***) (Table 2). From these partition coefficients, the relative free energies between every different state are calculated as $\Delta\Delta G = -RTlnQ$, where R is the gas constant, and T is the temperature of the experiment (298 K).

In addition to relative stabilities, the relative free energy barriers of each conformational transition may be approximated by relationship:

$$\Delta G^{t} = -RT\ln\left(\frac{rate \cdot h}{\kappa \cdot k \cdot T}\right),$$

where *rate* is the kinetic rate constant of the transition (sec^{-1}), h is Planck's constant, κ is the transmission coefficient, and k is Boltmann's constant (*6*). These free energy barriers are presented as relative barriers due to the inability to accurately determine the transmission coefficient (κ) for these biomolecular complexes (*6*) (Figure 7).

Similarity transform method to simulate evolution of states from the kinetic model

The similarity transform is a matrix method in which the probability distribution of states of an ensemble of MutS molecules incubated with GT-mismatched DNA, given the experimentally determined kinetic rates, may be calculated from first order differential equations. The procedure places the system of first order differential equations with constant coefficients (determined from each experimentally determined rate constant) into matrix form (A) and integrates the system, requiring a change of basis for matrices with nondiagonal elements. The products of each eigenvalue and the chosen

propagation time are exponentiated. The matrix is back-transformed by multiplication by the eigenvectors and then its inverse, and the resulting matrix is multiplied by the initial condition vector to yield the populations of each species at time, t (*7*).

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

$$\frac{d}{dt}x = (P^{-1}AP) * x(t)$$

$$x(t) = (P^{-1}\begin{bmatrix} e^{\lambda_1 * t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_1 * t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda n * t} \end{bmatrix} * P) * x(0)$$

$$A = \begin{bmatrix} -0.445 & 0.054 & 0.176 & 0.065 & 0.260 & 0.022 & 0.005 \\ 0.041 & -0.671 & 0.066 & 0.033 & 0.140 & 0.026 & 0.026 \\ 0.019 & 0.010 & -1.311 & 0 & 0.035 & 0.006 & 0.012 \\ 0.065 & 0.047 & 0 & -0.156 & 0.004 & 0.004 & 0.006 \\ 0.049 & 0.036 & 0.028 & 0.005 & -0.699 & 0.002 & 0.007 \\ 0.251 & 0.378 & 0.591 & 0.029 & 0.104 & -0.063 & 0.011 \\ 0.020 & 0.146 & 0.450 & 0.024 & 0.156 & 0.004 & -0.067 \end{bmatrix}$$

The output of this method may be viewed in two ways: (1) as the fraction of each unique species in an ensemble (Figure 8), or (2) as the likelihood of finding an individual DNA-MutS complex in a particular state at a certain time. Applying the similarity transform procedure to the experimentally determined kinetic mechanism for GT-MutS complexes including binding and unbinding events demonstrates conservation of species and provides the relative populations of each unique species as a function of time (Figure 8). An important constraint for these simulations is to verify that the kinetic scheme results in an equilibrium ensemble similar to that observed in the experiments.

The results demonstrate that the concentration of MutS affects both the evolution of the different conformational species and the time required for GT-MutS complexes to reach equilibrium (Figure 8A-B). Although the relative populations of the GT-MutS states are independent of MutS concentration (except there is a larger population of free DNA) when the system reaches equilibrium, as expected (Figure 8A-B), the simulations reveal differences in the evolution of the species as they approach equilibrium at 20 nM versus 200 nM MutS. Specifically, at 200 nM there is a transient increase in several species, especially ***B**** and ***I***, followed by a decrease that is not seen in the simulations for 20 nM (Figure 8A and 8B, inset). Similarly, the time to reach equilibrium as well as transient species evolved also depends on the initially bound state (data not shown).

Generation of Simulated Transition Density Plot

To generate the simulated transition density plot (shown in Fig. 2D), 3,000 representative transitions from Monte Carlo were extracted and assigned a start and stop FRET value. The values initially chosen for each state (based on the 2D Gaussian fits of the experimental data) were 0.26 for ***Free DNA***, 0.35 for ***U***, 0.47 for ***U*/I***, 0.57 for ***B****, 0.65 for ***B***, and 0.74 for ***SB***. The breadths of the distributions were initially modeled by adding or subtracting a random number with a standard deviation of 0.0075 to each start and stop FRET value. This procedure resulted in a transition density plot similar to that from experiment; however, the peaks in the plot were too circular in character and did not accurately represent the breadths of the distributions for the experimental data.

To generate a simulated TDP more representative of the experimental TDP, the breadths of the distributions were elongated along the diagonal of the TDP to account for

the heterogeneous broadening characteristic of single-molecule FRET trajectories (*4*). Instead of assigning the breadth of 0.0075 to the start and stop FRET values independently, two random numbers were added or subtracted to the FRET values. The first random number was generated and applied to each pair of start and stop FRET values to represent bias of transitions where the FRET of the states are both higher or lower than the average FRET of these states (systematic error). A second random number representing noise is generated for each start and stop FRET value separately. A series of plots were generated to determine the combination of bias and randomness that creates a TDP most similar to the experimental data. The breadths were varied individually over a wide range, and the simulated TDPs were compared to the experimental TDP (Figure 2A). The parameters found to best reproduce the experimental TDP were 0.36 for the bias parameter and 0.10 for the random error parameter (Figure 2D).

Supplemental Results and Discussion

Similarity transform method

The results from the similarity transform simulations were compared to calculations of probabilities based on experimental data (the product of the inverse of the lifetime of the state and the number of occurrences), and the total difference between the distributions differed by 5.5%. Because the error in the experimental rates did not result in perfect thermodynamic balance, minor adjustments were made in some of the experimentally determined kinetic rates to achieve thermodynamic balance between all states for the simulations. These adjustments are noted in Table S1. The simulated

ensemble using the original rates (Table 2) and the modified rates (Table S1) differed in placement of 3.7% of the events. But, by forcing thermodynamic balance, the modified parameters better represented the experimental data, and the total difference between the distributions was reduced from 5.5% to 3.4%.

Given a kinetic model for a biological system, the similarity transform method can be very helpful in both visualization of the evolution of species over time in addition to predicting the time it takes for a system to reach equilibrium given specific kinetic parameters. Starting with 100% free DNA, these simulations suggest that equilibrium is achieved in about 20 seconds at 200 nM MutS, but the system takes nearly 80 seconds to reach conformational equilibrium when the concentration of MutS is reduced to 20 nM (Figure 8).

Monte Carlo simulations validate the kinetic mechanism determined by FRET TACKLE

The similarity transform procedure does not explicitly yield the probability that a molecule will enter a state a given number of times in a given time window. Therefore, we performed Monte Carlo analysis to simulate each molecule's trajectory (states visited as a function of time (*8*)), which can be compared directly to the single-molecule FRET data.

Monte Carlo results were successful in validating the kinetic scheme in two ways. First, ensemble data from Monte Carlo agrees strongly with the similarity transform equilibrium ensemble data (total differences of 0.19% and 0.033% for the raw and modified parameters, respectively). Second, the probability of occurrence for each transition based on the experimental FRET data was computed and compared to the

respective probability of occurrence of that same transition in the Monte Carlo simulations (Table S2). Three sets of MC simulations were run: one containing 3,000 transitions (approximately the same number of transitions observed experimentally); one containing 30,000 transitions; and, one containing more than 60,000,000 transitions. Two thousand trials from the first two data sets with 3000 and 30,000 transitions were individually compared to the experimental data and to results from the third data set of more than 60,000,000 transitions (converged MC simulation). The data presented represent the average error of each of the 2000 trials (and not the trials combined) compared to converged or experimental data. This analysis was used to estimate both the stochasticity (by presenting a reduction of error with increased sampling) and the error in the kinetic scheme for GT-MutS complexes. The average absolute difference of 9.4% is reduced only to 6.0% when the MC simulation is compared to experimental data and includes 30,000 transitions instead of 3000, suggesting that 3000 transitions are sufficient to determine a kinetic scheme representative of the dynamics of the DNA-MutS complex. Comparison of the transition frequencies from the MC trials with 3000 and 30,000 transitions to those with 60,000,000 transitions (converged MC trial) yields absolute errors of 7.5% and 2.5% for trials of 3,000 and 30,000 transitions, respectively. The error between 3000 and 60,000,00 transitions generated by Monte Carlo (7.5%) is very similar to the error between the 3,000 MC generated transitions and the experimental data (9.4%) further indicating that the mechanism describes the data well.

Finally, comparison of the converged Monte Carlo data to the experimental data results in an error of 5.4%. In the Monte Carlo simulations, 3.7% of states have lifetimes of 0.1 seconds or less, which are faster than the rate of data collection and, therefore, are

not captured in the experiment. These missed states may partially contribute in the differences between the experimental and Monte Carlo transition data.

Monte Carlo simulations offer additional insight to the occupancies and preferred pathways of different conformational states in the MutS-GT mismatch recognition complex

Monte Carlo simulations of single binding events (MCSBE) were also performed and reveal important information about the occupancy of each of the different states for a given initial state (***U, U*, I, B*, B,*** or ***SB***) of the single-molecule simulation (Table S3) as well as the probability that a state will be visited at least once during a single binding event (Table 3). For each state on the DNA as the starting state for the simulation, 20,000 trials were run and the MCSBE simulation was halted when the ***Free DNA*** conformation is reached (*i.e.* MutS dissociates from the DNA), and this dissociation is given a relative occupancy value of 1.0 (Table S3). Transitions that are more likely to occur have high occupancy values, and those least likely to occur have lower values.

These results readily illustrate the dominance of the "productive" pathway (***U*↔*B***). With ***B*** as the initial state, the complexes will most likely occupy state ***U*** in a single binding event (1.055) compared to any other state, but have approximately a similar likelihood to dissociate (1.000) or to reoccupy state ***B*** (0.922) (Table S3). The productive pathway is even more pronounced with ***U*** as the initial state. The complexes not only have the highest chance of visiting the bent state ***B*** (1.486), but this number is the highest of all numbers representing the occupancies in MCSBE. Similar to simulations with ***B*** as the initial condition, MCSBE trials show a similar likelihood to

dissociate (1.000) or to reoccupy the initially bound state ***U*** (0.956) (Table S3). Initially binding in the intermediately bent state ***I*** is least likely to result in such transitions because it is most likely to result in dissociation of MutS from the DNA than to transition to any other state. In fact, dissociation from state ***I*** has a higher likelihood than the sum of all visits to all other states during a single binding event (Table S3). However, when excluding dissociation, transitions from ***I*** to ***B*** and ***I*** to ***U*** are observed more frequently than those from ***I*** to any of the other states (Table S3).

The remaining states (***U****, ***B****, and ***SB***) also show preference for the productive pathway compared to transitioning to any of the other bound states, but the relationships among these states to each other are more surprising as they show tendencies to transition to one another. The highest likelihood of forming the unstable unbent and unstable bent states (***U**** and ***B****) occurs when ***SB*** serves as the initial state of the simulation (relative occupancies of 0.427 and 0.284, respectively). Conversely, the highest likelihood of forming the super-bent state ***SB*** occurs when the initial states are ***U***, ***U****, or ***B****. These results suggest that ***U****, ***B****, and ***SB*** may comprise a subnetwork of states that support the productive pathway.

The Monte Carlo results, combined with the experimentally-identified kinetic mechanism of the MutS-GT complex, are fundamental in understanding the kinetic properties of each of the states and are important in the interpretation of the roles of each of these states in the MutS-mismatch initiation complex.

Power and Limitations of FRET TACKLE analysis

Studies using single-molecule FRET are increasing at an exponential rate due to this method's unique capability to detect the conformational dynamics and fluctuations of individual biomolecules (*9, 10*). Much of the recent work using single-molecule FRET has focused on studying motor proteins, conformational changes or folding in single proteins or nucleic acids, or binding interactions between biomolecules (*11-21*). Here we report single-molecule FRET studies of the spontaneous, equilibrium conformational dynamics of individual DNA-protein complexes. As single-molecule FRET studies become more sophisticated and are extended to more dynamic biological systems, such as the one described here, the resulting data become exceedingly complex, generating a need for additional analysis approaches to identify unique states and kinetic rates of exchange from dynamic FRET traces. Current analysis methods are powerful in teasing apart dynamic FRET distributions but are limited by certain experimental factors (*3, 4, 22*). Determining transitions based on hidden Markov modeling, for example, requires that every conformational state be sampled in a single FRET trace (*4*). Because the dynamics and kinetic rates of exchange for some mismatched DNA-MutS complexes are slow relative to the lifetime of the fluorophores, it is difficult to meet this sampling requirement. In fact, many molecules only undergo one or two transitions within the few minutes typical of a single-molecule FRET trace (Figure S1). FRET TACKLE may be applied to a large population of molecules without being limited by the frequency with which a particular state occurs during a single trace (*4*).

In addition, previously developed methods on their own are insufficient in analyzing complex FRET efficiency distributions that result from a series of overlapping states. Kinetic fingerprinting has been applied to identify overlapping FRET efficiency

states, where exchanges of buffer components may be used to isolate and identify substeps in a well-defined enzymatic pathway (*23*). However, this approach is difficult to apply to equilibrium DNA-protein complex dynamics that do not exhibit a defined reaction trajectory. FRET TACKLE marries previously developed methods to allow detailed analysis of FRET efficiency distributions resulting from a system with a broad distribution of states and diverse dynamics. By combining analysis of both transitions and lifetimes, transition sequences for unique species residing in a single FRET efficiency state may be isolated.

Supplementing the FRET TACKLE method with the similarity transform procedure and Monte Carlo simulations of single-molecule trajectories confirms the self-consistency within the experimentally determined kinetic model and the relative thermodynamic stabilities of states identified by FRET TACKLE analysis. These simulation approaches also independently reveal the equilibrium distribution of states in a dynamic biological system, the evolution of biological species in a dynamic system, and the ability of our mechanism to reproduce the experimental results. Performing the similarity transform procedure at two different protein concentrations, for example, reveals differences in dynamic behavior of different species (Figure 8). While we apply the similarity transform procedure and Monte Carlo simulations to verify and complement the FRET TACKLE analysis, they may be extended to understanding the dynamics and evolution of conformational species in a host of other dynamic biological systems.

Like most analysis methods, the FRET TACKLE approach does have some limitations. First, the Gaussian derivative kernel edge analysis algorithm (Figure S2) is

highly dependent on selection of a threshold parameter. As a result, this approach is not as sensitive to transitions between states with closely spaced but distinct FRET efficiencies as is analysis using hidden Markov modeling (*4, 24, 25*). Additionally, the Gaussian derivative kernel approach may only detect transitions that are greater than the noise fluctuations in the FRET trace. Therefore, the ability to isolate a conformational transition is also limited by the noise in the FRET efficiency measurements. Transitions resulting in changes in FRET greater than ~ 0.10 are detected, but transitions resulting in changes in FRET less than ~ 0.10 (FRET transition of 0.5 to 0.56, for example) are sometimes overlooked by this analysis simply as noise in the trace. Furthermore, FRET TACKLE does not allow us to detect transitions between different states with identical average FRET efficiencies (such as between state ***U**** and ***I***).

Finally, fast conformational transitions, such as those lasting for fewer than 5 frames (0.5 seconds in these experiments), are often missed in the Gaussian derivative kernel edge analysis algorithm due to smoothing (Figure S2, green arrow). Of more than 3000 states sampled in the experiment, less than 1% of them had lifetimes shorter than 0.5 seconds, revealing that only a very small fraction of these transitions were experimentally captured. Analysis of lifetimes obtained from Monte Carlo simulations of the kinetic mechanism reveals that 3.7% of lifetimes are less than 0.1 seconds and 16.1% of lifetimes, less than 0.5 seconds. This limitation of the Gaussian derivative kernel edge analysis is, therefore, partially overcome by supplementing the experimental results with Monte Carlo simulations.

A future combination of FRET TACKLE with other edge analysis and state identification algorithms (such as hidden Markov modeling, for example) may prove

effective in analyzing complex single-molecule FRET data as the advantages of one method overcome the limitations of the other.

Supplemental Figure Legends

Figure S1. Single-molecule FRET traces for GT-mismatched DNA in the presence of MutS on two surfaces: a BSA-biotin-streptavidin linkage (A, represented in Figure 1), and polyethylene glycol (PEG)-biotin-streptavidin linkage (B). Results show similar GT-MutS dynamics on both surfaces, revealing that these dynamics are not the result of interactions of the DNA or the protein with the surface. Kinetic rates of exchange among different states in these complexes varied, and some molecules underwent only a single transition during the FRET time trace, limited by bleaching of the fluorescent dye (shown in Figure 1A).

Figure S2. An example time trace showing the fluctuations in the donor fluorescence intensity and the acceptor fluorescence intensity (green and red traces, respectively, (a)) and the calculated FRET efficiency from these corresponding traces (black trace, (b)). The derivative of Gaussian kernel (FRET') to detect transition edges in this trace is shown in (c). Blue tick marks represent extrema in the derivative of Gaussian kernel. From this approach, unique conformational states (numbered in (b)), the dwell times at each of these states (Δt), and the transition sequence among different states (magenta arrows in (b)) may be identified.

Figure S3. Dwell time distributions for conformations ***B*** and ***SB*** separated by FRET efficiency. The dwell time distribution for conformation ***B*** (A) includes species from the mixed FRET state at FRET 0.51 to 0.60 in addition to species residing at FRET 0.61 to 0.70. The dwell time distribution for conformation ***SB*** (B) represents dwell times

associated with all FRET efficiencies from 0.71 to 1.0. As expected from the distributions shown in Figure 3, the lifetimes resulting from the separate fits of these two states converge to similar values, confirming the inability to identify these two conformational states by lifetime analysis alone.

Figure S4. Transition density slices of individual FRET efficiency states. Each slice is shown (a), as well as the cross section of each slice (b-f), showing every transition among different FRET efficiency states across the entire transition density distribution (Figures 2 and 4). Mixed FRET efficiency states (U* + I and B* + B) were further separated by lifetime (analysis shown in Figures 5 and S5).

Figure S5. (A) and (B) Dwell time distributions of FRET efficiency states comprised of two species determined from lifetime analysis ('mixed FRET states'). The double exponential decay fit (solid black line) is deconvoluted into two single exponential decay fits (dashed black lines), and the two species are separated by dwell time. (C) through (F) Transition density distributions for transitions observed when the two species in the mixed FRET states are separated by dwell time. Scatter plots represent additional cross-sections of each transition. The transition sequences of each species are not affected by altering the cutoff time between the two states to +/- 1 second.

Figure S6. Sample traces of single molecules from donor-only and acceptor-only control experiments in the absence or presence of MutS. Also shown are additional traces showing a direct comparison of homoduplex DNA in the absence or presence of MutS, as

well as a side-by-side comparison of GT mismatched DNA in the absence or presence of MutS on either a biotinylated-PEG surface or a biotinylated-BSA surface.

		Transition rate (s^{-1})	$N_{x \to y} / N_{x,total}$	$Q = k_f / k_r$
GT-MutS complex transitions	*U to U**	*0.01**	0.014	*0.15**
	U to U*	*0.066**	0.059	
	U to I	0.047	0.07	*1.4**
	I to U	*0.033**	0.12	
	*U to B**	0.036	0.053	*0.26**
	B to U*	*0.14**	0.25	
	U to B	0.38	0.56	15
	B to U	0.026	0.42	
	U to SB	0.15	0.22	*5.8**
	SB to U	*0.026**	0.31	
	U to B*	0.59	0.45	95
	*B to U**	0.0062	0.099	
	U to SB*	0.45	0.34	38
	*SB to U**	0.012	0.17	
	I to B	*0.029**	0.25	*11**
	B to I	0.0027	0.043	
	I to SB	*0.024**	0.18	*3.9**
	SB to I	0.0062	0.087	
	B to B*	0.10	0.15	53
	*B to B**	0.0019	0.030	
	B to SB*	0.16	0.23	21
	*SB to B**	0.0075	0.10	
	B to SB	0.0039	0.063	*0.35**
	SB to B	*0.011**	0.19	
				K_A (M^{-1})
MutS Binding/ Unbinding transitions	*free to U*	0.041	0.096	3.8×10^6
	U to free	0.054	0.081	
	*free to U**	0.019	0.043	5.3×10^5
	U to free*	0.18	0.13	
	free to I	*0.065**	0.12	*5.0×10^6 **
	I to free	0.065	0.42	
	*free to B**	0.049	0.11	*9.4×10^5 **
	B to free*	*0.26**	0.31	
	free to B	0.25	0.58	5.7×10^7
	B to free	0.022	0.35	
	free to SB	0.020	0.046	*2.0×10^7 **
	SB to free	*0.005**	0.14	

Table S1. Kinetic rates, transition probability ratios, and partition ratios of GT-MutS binding, unbinding, and complex transitions. The transition rates that were modified to induce thermodynamic balance for the simulations of the kinetic model are marked with asterisks.

A

State after transition

State before transition / *FRET*	***Free DNA***	***U***	***U****	***I***	***B****	***B***	***SB***
Free DNA	-	0.0197	0.0088	0.0248	0.0235	0.1196	0.0095
U	0.0177	-	0.0031	0.0153	0.0116	0.1227	0.0476
U*	0.0085	0.0037	-	*nd*	0.0014	0.0286	0.0218
I	0.0207	0.0061	*nd*	-	0.0017	0.0122	0.0088
B*	0.0163	0.0133	0.0027	*nd*	-	0.0082	0.0122
B	0.1020	0.1220	0.0289	0.0126	0.0088	-	0.0184
SB	0.0160	0.0364	0.0197	0.0102	0.0122	0.0228	-

B

State after transition

State before transition / *MC*	***Free DNA***	***U***	***U****	***I***	***B****	***B***	***SB***
Free DNA	-	0.0182	0.0082	0.0229	0.0216	0.1104	0.0088
U	0.0167	-	0.0029	0.0145	0.0110	0.1165	0.0452
U*	0.0085	0.0037	-	*nd*	0.0013	0.0285	0.0217
I	0.0254	0.0075	*nd*	-	0.0021	0.0149	0.0108
B*	0.0178	0.0145	0.0030	*nd*	-	0.0089	0.0134
B	0.1055	0.1262	0.0298	0.0130	0.0091	-	0.0188
SB	0.0162	0.0368	0.0200	0.0103	0.0124	0.0231	-

C

State after transition

State before transition / *% difference*	***Free DNA***	***U***	***U****	***I***	***B****	***B***	***SB***
Free DNA	-	0.1514	0.0638	0.1928	0.1878	0.9216	0.0700
U	0.0968	-	0.0193	0.0757	0.0600	0.6209	0.2425
U*	0.0025	-0.0040	-	*nd*	0.0053	0.0076	0.0055
I	-0.4652	-0.1379	*nd*	-	-0.0389	-0.2738	-0.2017
B*	-0.1527	-0.1179	-0.0267	*nd*	-	-0.0723	-0.1164
B	-0.3458	-0.4155	-0.0937	-0.0402	-0.0349	-	-0.0377
SB	-0.0179	-0.0408	-0.0259	-0.006	-0.0156	-0.0322	-

Table S2. Comparison of experimentally observed transitions between states (A) and Monte Carlo simulations of transitions between states based on the experimentally determined kinetic model (B).

(A) Transitions calculated between states by FRET TRACKLE from single-molecule FRET traces (shown as a fraction of the total number of transitions (n = 2942)).

(B) Transitions calculated between states by Monte Carlo simulations shown as a fraction of the total number of transitions simulated (n = 60,096,000).

(C) Calculated difference between the experimentally determined fraction of transitions (A) and the fraction of transitions observed in Monte Carlo simulations (B) (shown as absolute percent differences). Summing the absolute values of this matrix's entries reveals that the total % error of misassigned transitions is 5.4%

A

Occupancy of states during MCSBE simulation	*MCSBE initial state*					
	U	***U****	***I***	***B****	***B***	***SB***
Free DNA	*1.000*	*1.000*	**1.000**	***1.000***	***1.000***	1.000
U	*0.956*	*0.987*	0.720	***0.978***	**1.055**	***1.143***
U*	0.301	0.308	0.190	0.295	0.302	0.427
I	0.263	0.201	0.132	0.173	0.204	0.267
B*	0.241	0.219	0.161	0.167	0.182	0.284
B	**1.486**	**1.380**	***0.889***	**1.042**	***0.922***	**1.254**
SB	0.724	0.842	0.509	0.681	0.530	0.588
Total # transitions	4.97	4.94	3.60	4.34	4.19	4.96

B

	MCSBE initial state						
	U	***U****	***I***	***B****	***B***	***SB***	***Average at equilibrium***
Residence time on the DNA	38.80	37.81	29.82	30.33	41.44	46.50	38.61
# unique states sampled in SBE (excluding initial state unless it was revisited)	2.40	2.36	1.57	1.99	1.91	2.42	1.97
# unique states sampled in SBE (including the initial state)	2.91	3.12	2.45	2.85	2.43	3.05	2.59

Table S3. (A) The relative occupancies of different states (denoted in the leftside column) given a particular starting conformational state (denoted in the top row) in a Monte Carlo single binding event (MCSBE) simulation. The simulation was halted when MutS dissociated (or, the ***Free DNA*** conformation was reached). As a result, the occupancy of states is normalized and compared to that of ***Free DNA***, which is set = 1.0. Larger occupancy values represent transitions that are most likely to occur, and smaller

occupancy values represent transitions that are least likely to occur. These results show, for example, that given the GT-MutS molecules start in state ***B*** (top row, column ***B*** reading down), state ***U*** would be the most frequently occupied (relative occupancy of 1.055). These results also reveal which starting state would produce the greatest likelihood for a given state to be occupied. For example, the occupancy of state ***B**** (leftside column, row ***B**** reading right) is highest given the GT-MutS molecules start in state ***SB*** (relative occupancy of 0.284). Values in bold represent the transitions that are most likely to occur for each starting state in a MCSBE simulation, with those in bold-italics representing other transitions with a high likelihood of occurring in the simulation. (B) The number of states visited in MCSBE for a given starting state in the simulation, both including and excluding the initial state, and the average number of states visited at equilibrium.

References

(1) Ha, T., Rasnik, I., Cheng, W., Babcock, H. P., Gauss, G. H., Lohman, T. M., and Chu, S. (2002) Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature 419*, 638-641.

(2) Canny, J. (1986) A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell. 8*, 679-698.

(3) Zhuang, X., Kim, H., Pereira, M. J., Babcock, H. P., Walter, N. G., and Chu, S. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science 296*, 1473-1476.

(4) McKinney, S. A., Joo, C., and Ha, T. (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys J 91*, 1941-1951.

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

(6) Atkins, P. W. (1978) *Physical Chemistry*, Oxford University Press.

(7) Davis, J. H. (2001) *Differential Equations with Maple: An interactive approach*, Birkhauser.

(8) Gillespie, D. T. (1976) A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *Journal of Computational Physics 22*, 403-434.

(9) Moerner, W. E. (2007) New directions in single-molecule imaging and analysis. *Proc Natl Acad Sci U S A 104*, 12596-12602.

(10) Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C., and Ha, T. (2008) Advances in single-molecule fluorescence methods for molecular biology. *Annu Rev Biochem 77*, 51-76.

(11) Deniz, A. A., Laurence, T. A., Beligere, G. S., Dahan, M., Martin, A. B., Chemla, D. S., Dawson, P. E., Schultz, P. G., and Weiss, S. (2000) Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc Natl Acad Sci U S A 97*, 5179-5184.

(12) Lipman, E. A., Schuler, B., Bakajin, O., and Eaton, W. A. (2003) Single-molecule measurement of protein folding kinetics. *Science 301*, 1233-1235.

(13) Margittai, M., Widengren, J., Schweinberger, E., Schroder, G. F., Felekyan, S., Haustein, E., Konig, M., Fasshauer, D., Grubmuller, H., Jahn, R., and Seidel, C.

A. (2003) Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. *Proc Natl Acad Sci U S A 100*, 15516-15521.

(14) Rhoades, E., Cohen, M., Schuler, B., and Haran, G. (2004) Two-state folding observed in individual protein molecules. *J Am Chem Soc 126*, 14686-14687.

(15) Rhoades, E., Gussakovsky, E., and Haran, G. (2003) Watching proteins fold one molecule at a time. *Proc Natl Acad Sci U S A 100*, 3197-3202.

(16) Weninger, K., Bowen, M. E., Chu, S., and Brunger, A. T. (2003) Single-molecule studies of SNARE complex assembly reveal parallel and antiparallel configurations. *Proc Natl Acad Sci U S A 100*, 14800-14805.

(17) Myong, S., Rasnik, I., Joo, C., Lohman, T. M., and Ha, T. (2005) Repetitive shuttling of a motor protein on DNA. *Nature 437*, 1321-1325.

(18) Li, Y., Augustine, G., and Weninger, K. (2007) Kinetics of complexin binding to the SNARE complex: correcting single molecule FRET measurements for hidden events. *Biophys J*.

(19) Kim, H. D., Nienhaus, G. U., Ha, T., Orr, J. W., Williamson, J. R., and Chu, S. (2002) Mg2+-dependent conformational change of RNA studied by fluorescence correlation and FRET on immobilized single molecules. *Proc Natl Acad Sci U S A 99*, 4284-4289.

(20) Ha, T. (2004) Structural dynamics and processing of nucleic acids revealed by single-molecule spectroscopy. *Biochemistry 43*, 4055-4063.

(21) Bowen, M. E., Weninger, K., Ernst, J., Chu, S., and Brunger, A. T. (2005) Single-molecule studies of synaptotagmin and complexin binding to the SNARE complex. *Biophys J 89*, 690-702.

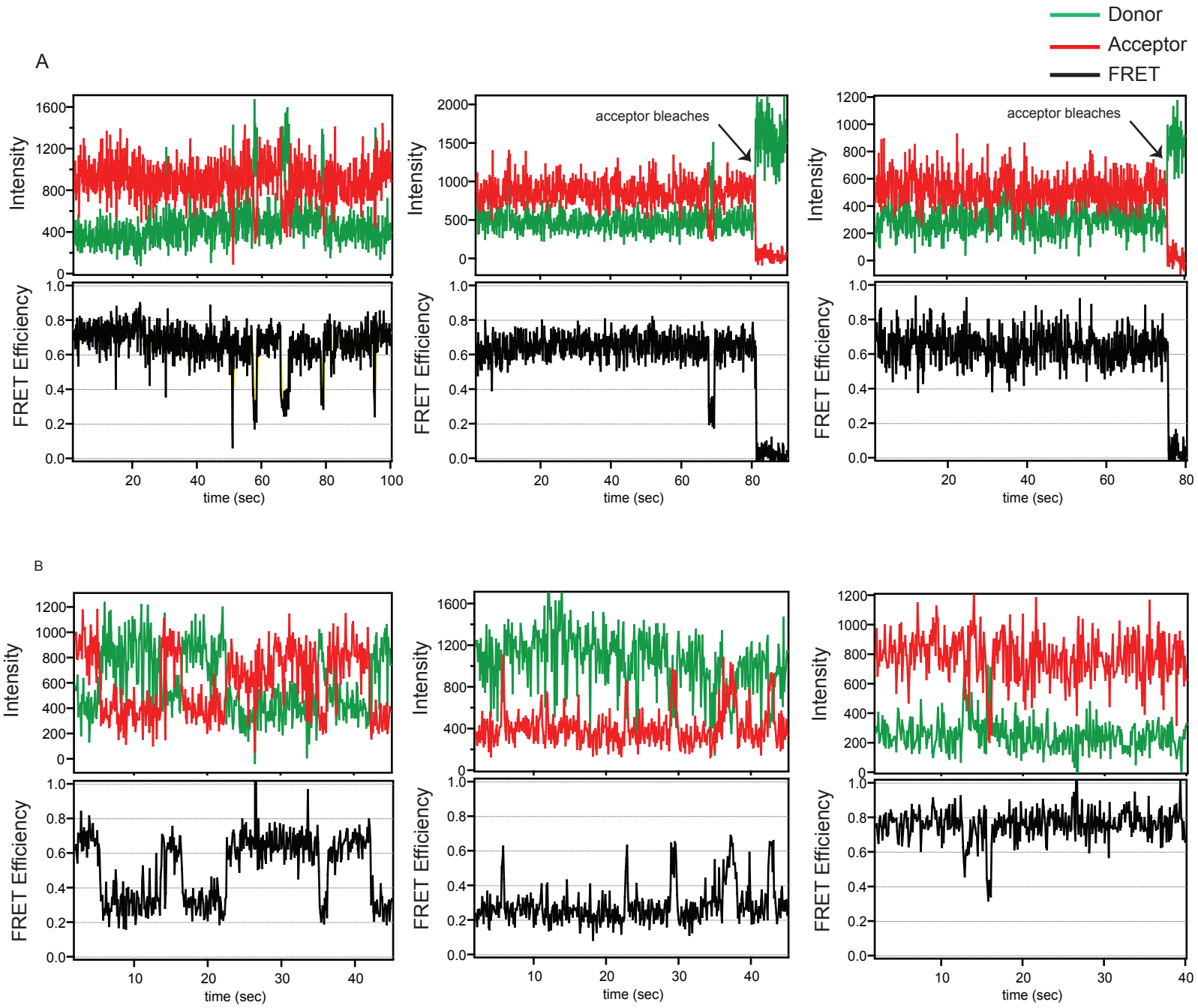
(22) Gell, C., Brockwell, D., and Smith, A. (2006) *Handbook of Single Molecule Fluorescence Spectroscopy*, Oxford University Press, New York.

(23) Liu, S., Bokinsky, G., Walter, N. G., and Zhuang, X. (2007) Dissecting the multistep reaction pathway of an RNA enzyme by single-molecule kinetic "fingerprinting". *Proc Natl Acad Sci U S A 104*, 12634-12639.

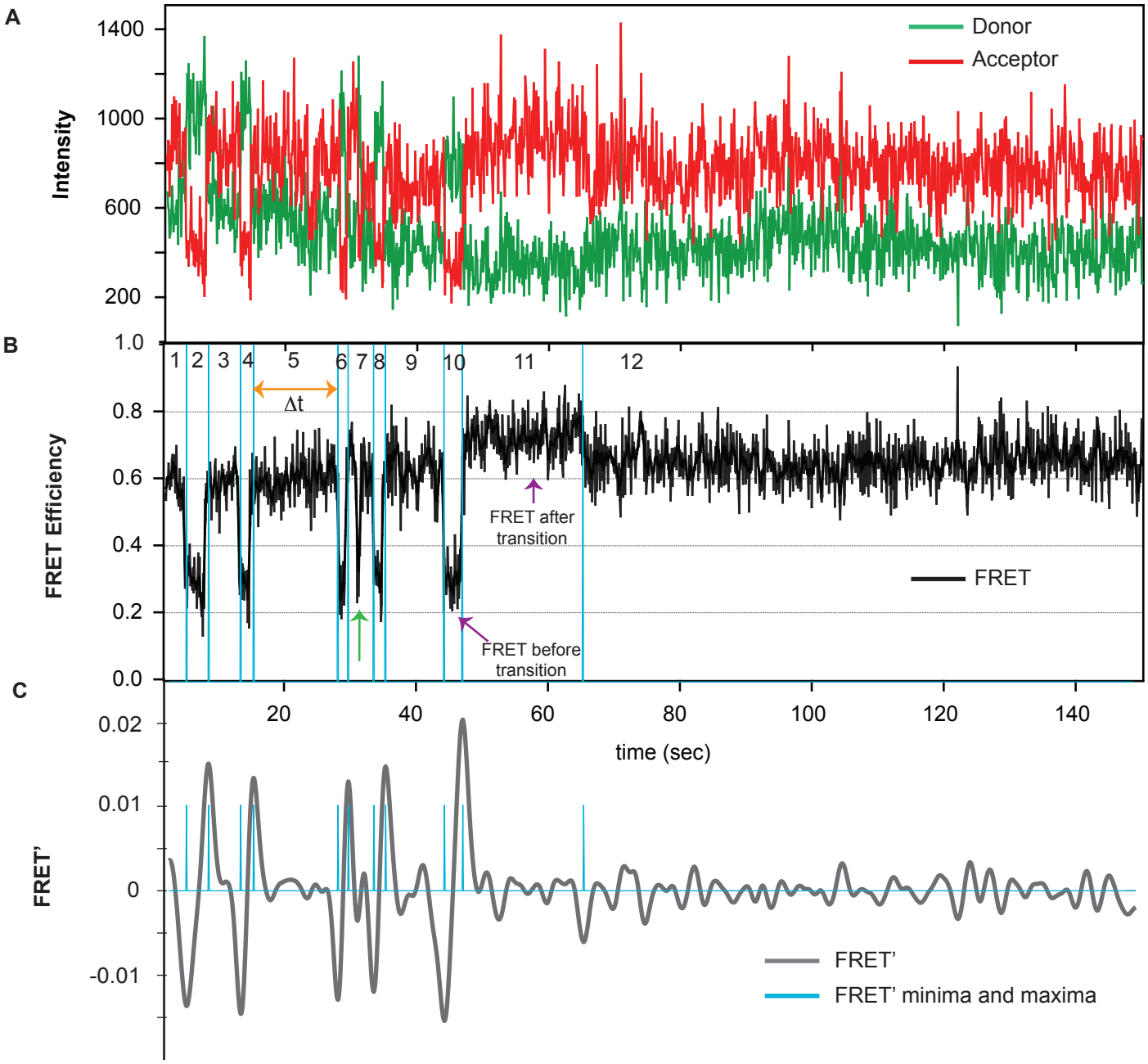
(24) Beausang, J. F., and Nelson, P. C. (2007) Diffusive hidden Markov model characterization of DNA looping dynamics in tethered particle experiments. *Phys Biol 4*, 205-219.

(25) Beausang, J. F., Zurla, C., Manzo, C., Dunlap, D., Finzi, L., and Nelson, P. C. (2007) DNA looping kinetics analyzed using diffusive hidden Markov model. *Biophys J 92*, L64-66.

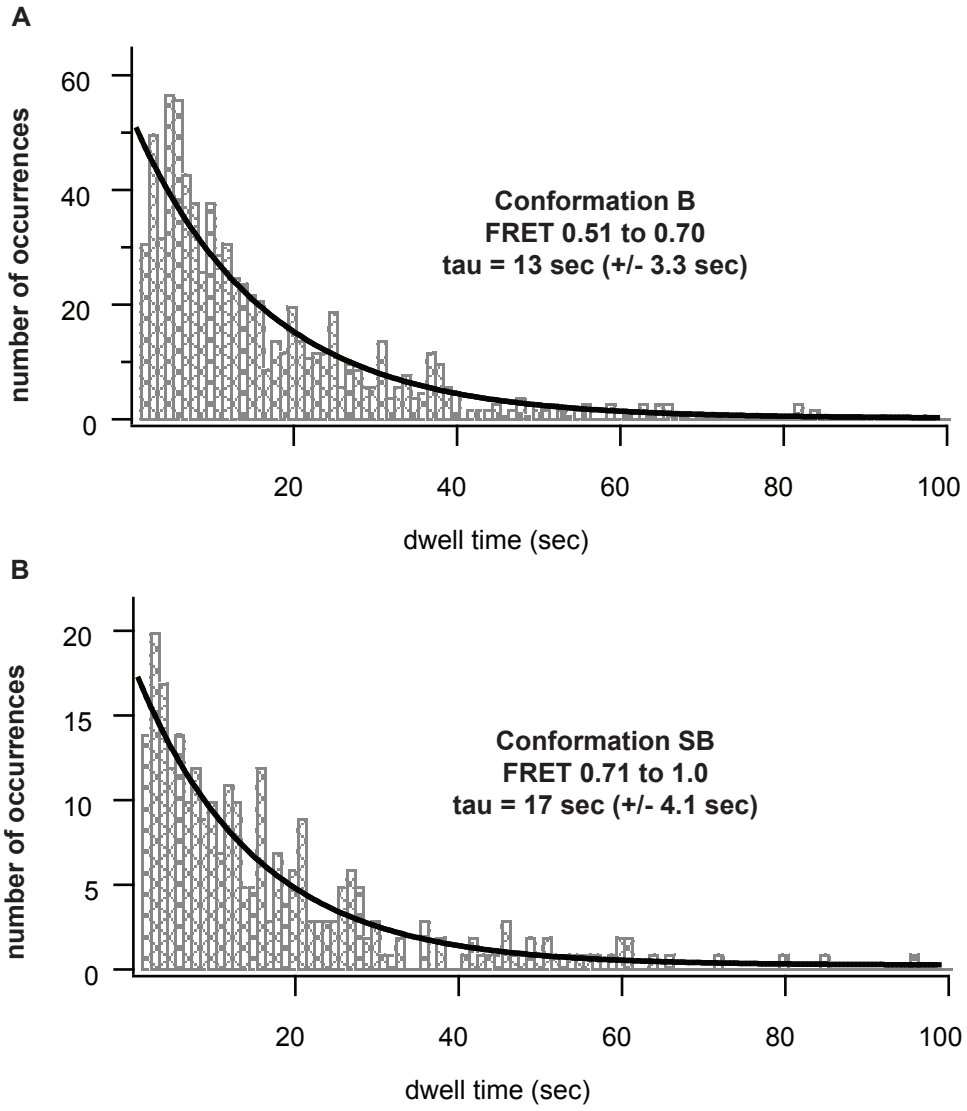
SASS FIGURE S1



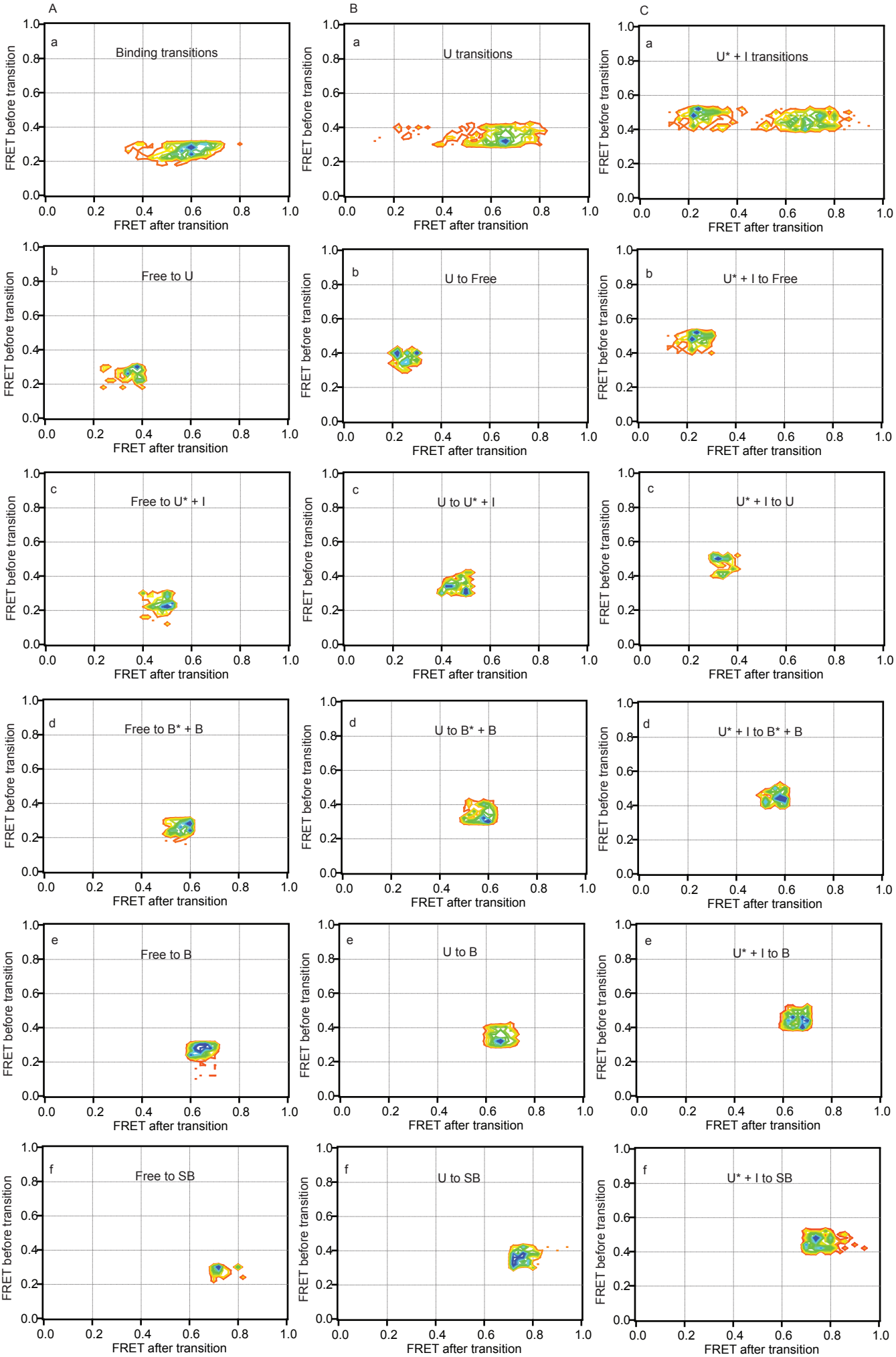
SASS FIGURE S2



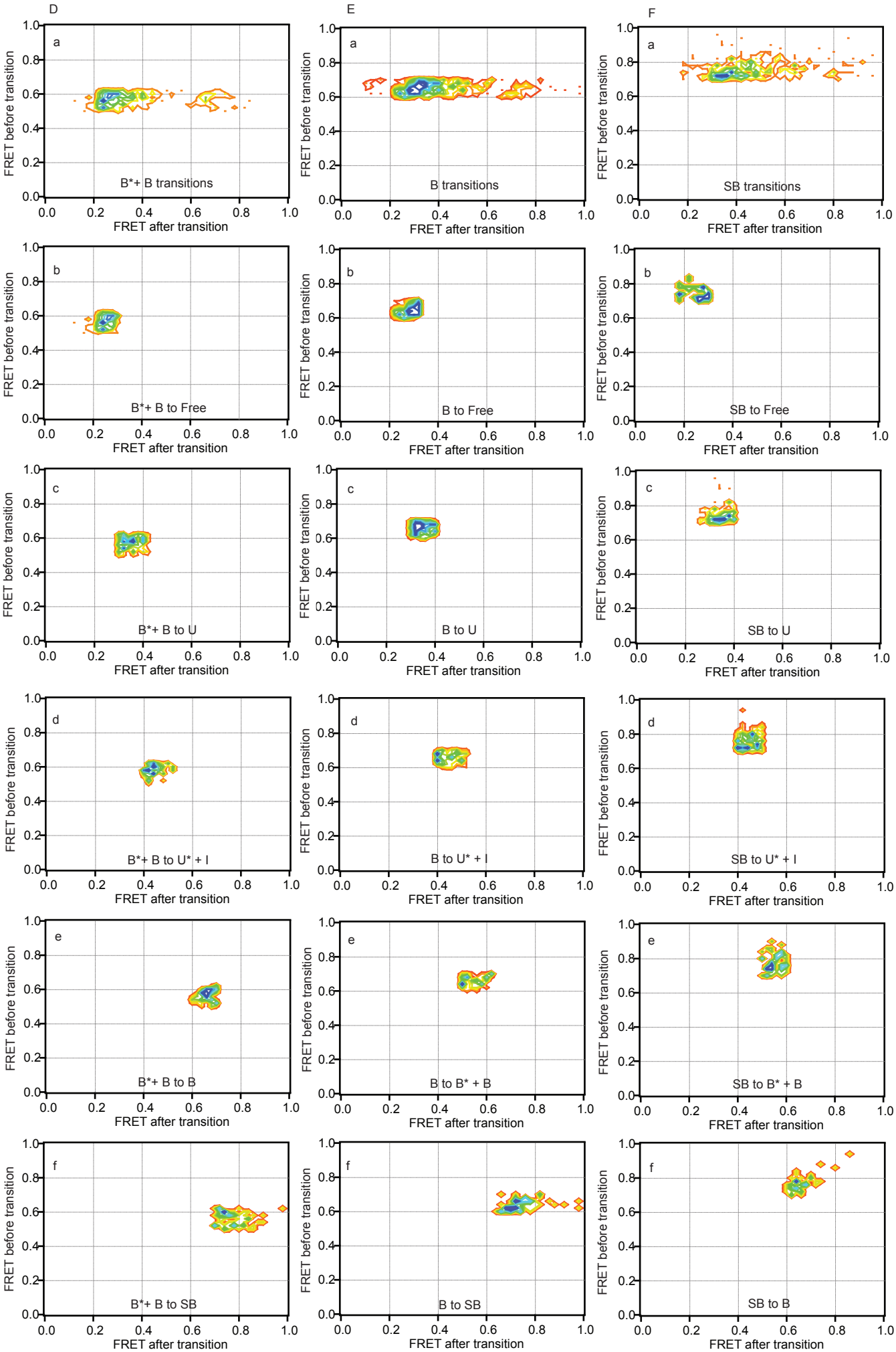
SASS FIGURE S3



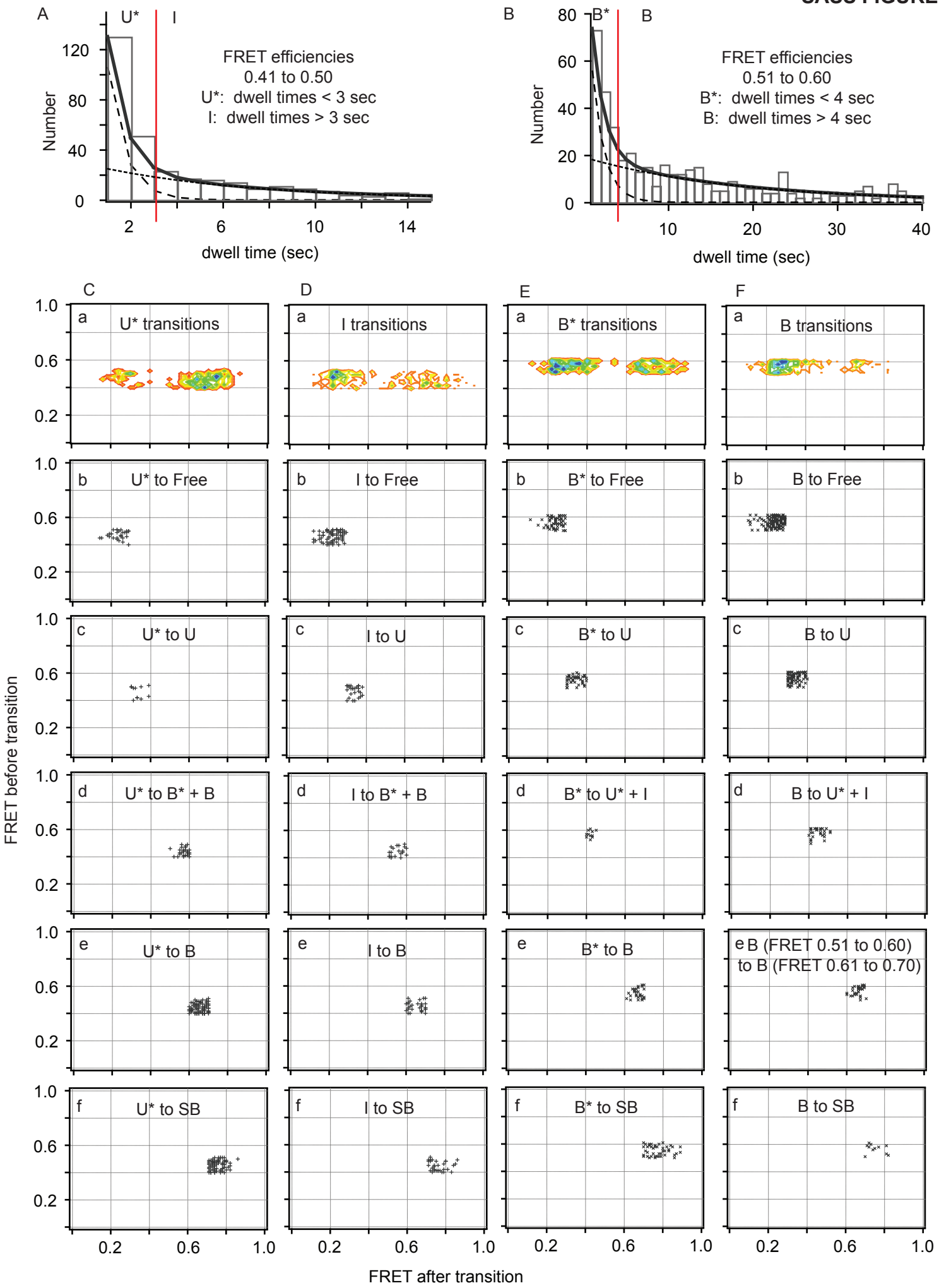
SASS FIGURE S4



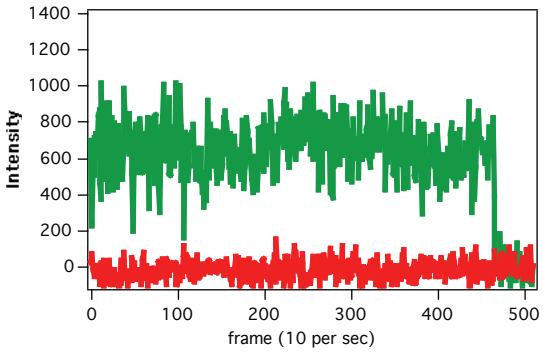
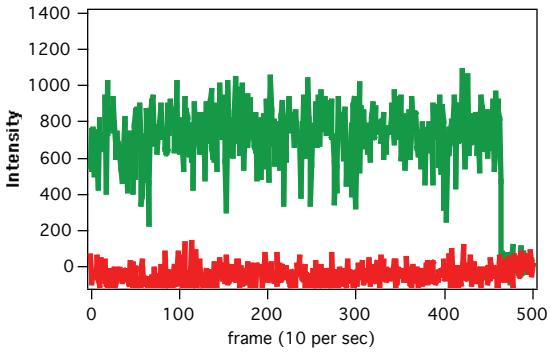
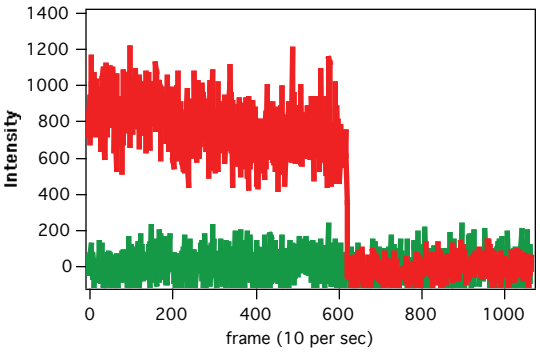
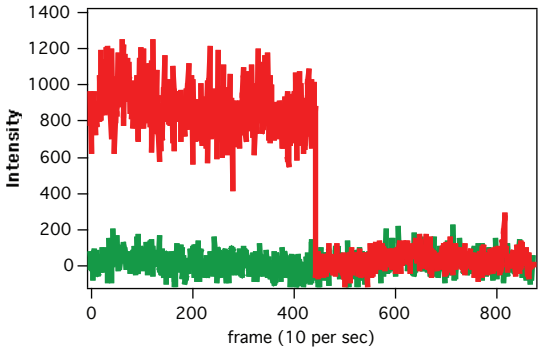
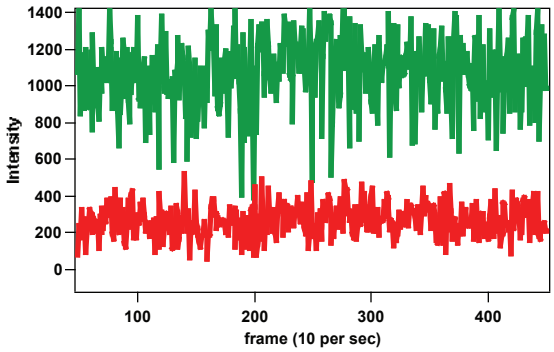
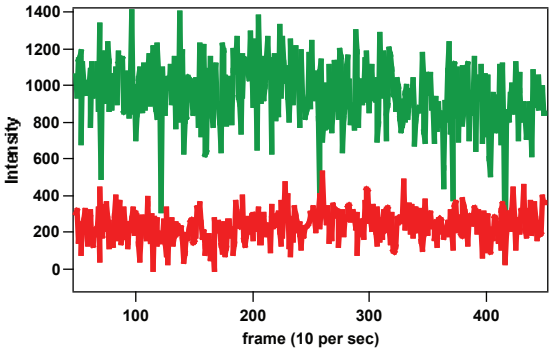
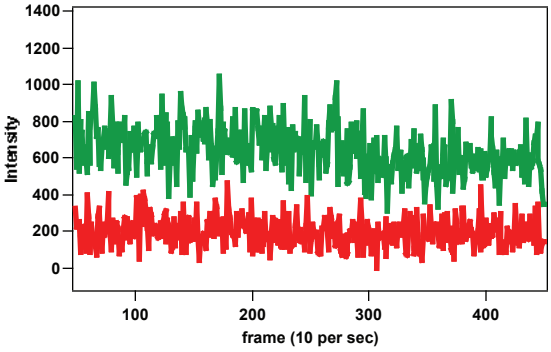
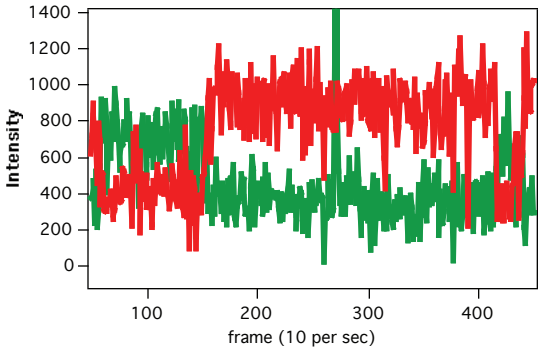
SASS FIGURE S4, continued



SASS FIGURE S5



SASS FIGURE S6

	- MutS	+ 200 nM MutS
Traces for Donor-only GT mismatched DNA (Green excitation)		
Traces for Acceptor-only GT mismatched DNA (Red direct excitation)		
Traces for Dual-labeled Homoduplex DNA		
Traces for Dual-labeled GT mismatched DNA PEG surface		
Traces for Dual-labeled GT mismatched DNA BSA surface