Supplementary Data

1 Supplementary Materials and Methods

1.1 Single-molecule spectroscopy

Single molecule measurements of a very dilute solution (~15 pM) of double-labeled DNA molecules were performed with a confocal epi-illuminated microscope. The donor fluorophore (Alexa 488) was excited by a linearly polarized, active-mode-locked Argon-ion laser (476.5 nm, 73 MHz, 300 ps). The laser light is focused into the solution with a 60x/1.2 water immersion objective. Fluorescence bursts arising by single molecules diffusing through the detection volume are detected. This photon-train is divided initially into its parallel and perpendicular components via a polarizing beamsplitter and then into wavelength ranges below and above 595 nm. Additionally, red (HQ 630/66 nm) and green (HQ 533/46 nm) filters in front of the detectors ensure that only photons emitted by the acceptor (Alexa 594) and donor (Alexa 488) molecules are registered. The focal geometry is estimated from the measured diffusion correlation time (200 ± 13 µs) and the known diffusion coefficient (0.34 ± 0.03 µm²/ms) for Rhodamine 110. In addition, correction factors \( l_1 = 0.0308 \) and \( l_2 = 0.0368 \) are used to account for the mixing of polarization by the microscope objective, and a factor \( G = 1.02 \) is applied to compensate for the slightly different detection efficiency of the two polarization components. Detection is performed using four avalanche photodiodes (SPCM-AQR-14, Laser Components, Germany). The signals from all detectors are passed through a passive delay unit and two routers to two synchronized time-correlated single photon counting boards (SPC 132, Becker and Hickl, Germany) that are connected to a PC. Fluorescence bursts are distinguished from the background of 1-3 kHz by applying a threshold intensity criterion (~0.05 ms interphoton time, 100 photons minimum per burst).

1.2 Modeling the 42bp G:T heteroduplex DNA

A 3D structure of the unbent DNA was obtained by introducing the sequence into the 3D-DART webserver (26). A model of the MutS-bound kinked DNA was obtained by using the DNA
distortion parameters obtained from the NDB server (ndbserver.rutgers.edu) using the crystal structure of MutS and DNA (PDB code 1e3m). Parameters of positions 2-16 from chain E (15-29 from chain F) were used for position 14-28 (57-71) of the 42bp G:T-DNA or positions 57-71 (14-28) for the T:G DNA, respectively, while using B-DNA parameters (obtained from the 3D-DART webserver using the same sequence) for all other positions. Sequence-dependent structural bending of the chosen DNA sequence is not significant as judged by the program DIAMOD (27.) MutS was docked to this modeled DNA by superposition of the backbone atoms of the DNA in the co-crystal structure onto the respective atoms of the modeled DNA using the Swiss PDB viewer (version 4.07).

1.3 Modeling the fluorophore position cloud obtained from MD simulations to the MutS-DNA complex structure.

We used accessible volume (AV) simulations to describe the rotational freedom of the fluorophore dependent on the linker length and steric clashes with the DNA (28). For each calculation of a position distribution, we used the real physical dimensions of the fluorophore and performed three independent AV simulations with three different radii $R_{\text{dye}(i)}$ and superimposed them. Thus, the obtained position distribution represents an average weighted by the number of allowed positions. Throughout this work we used for Alexa488 $R_{\text{dye}(1)} = 5$ Å, $R_{\text{dye}(2)} = 4.5$ Å and $R_{\text{dye}(3)} = 1.5$ Å and for Alexa594 $R_{\text{dye}(1)} = 6.7$ Å, $R_{\text{dye}(2)} = 4.5$ Å and $R_{\text{dye}(3)} = 1.5$ Å. It turned out that these “mixed” AV simulations are necessary to accurately predict dye distributions. In these simulation the length $l$ and the width $w$ of dyes linkers was the same for both dyes ($l = 20$ Å, $w = 4.5$ Å). The theoretical distance between the fluorophores was measured between the center of mass of the fluorophores’ accessible space. A Förster distance ($R_0$) of 53.2 Å was used to calculate the transfer efficiency $E$. For free B-DNA$\text{DNA}_{\text{DA}}$ the theoretical distance was calculated from the mean position of the fluorophores to be $R_{\text{mp}} = 85$ Å ($E_{\text{mp}}$ of 0.063).

2 Supporting theory for FRET analysis

2.1 Burst selection

As discrete molecules are diffusing through the detection volume, bursts of fluorescence photons are registered in the detected photon trace. These bursts are selected against background as
described in (58). In this study, additional burst selection steps were performed on the fluorescent bursts recovered from smMFD measurements, in order to reduce the background (FRET-inactive species) and allow the analysis to be constrained to the FRET-active subpopulations. Two selection steps were performed:

a) $-0.45 < T_g - T_r < 0.45$: Here the bursts during which an event of acceptor photobleaching occurs were excluded. The time difference $T_g - T_r$ of the mean observation time of all photons detected for the donor and acceptor channels within a single molecule fluorescence burst enables identification and exclusion of photobleached molecules (59). $T_g$ and $T_r$ represent the medial time of fluorescence emission of both detection channels. Without photobleaching, $T_g$ and $T_r$ would be similar and thus $T_g - T_r \approx 0$.

b) $N_{ph,RED} > 20$: Bursts with less than 20 photons in the red channel, $N_{ph,RED}$, were excluded from the analysis. This step enables exclusion of small red bursts that are not originated by FRET, but either by direct excitation of the acceptor or by crosstalk from green donor signal into the red acceptor detection channel. In this way, a significant fraction of the donor-only species was removed from the analysis.

The performed burst selection does not change the position or the widths of the FRET populations, and can be thus used to significantly reduce the background and facilitate the analysis.

### 2.2 Calculation of FRET and fluorescence anisotropy parameters

The efficiency, $E$, of FRET depends strongly on the interdye distance $R_{DA}$ and the Förster radius, $R_0$ (Equation 1).

$$
E = \left[ 1 + \left( \frac{R_{DA}}{R_0} \right)^2 \right]^{-1}
$$

(Equation 1)
Each fluorophore pair has a characteristic Förster radius, \( R_0 \), which accounts for the photophysical properties and relative orientation of the dyes. It is calculated (in Å) by

\[
R_0 = \left( c_{FT} J(\lambda) \kappa^2 \Phi_{FD(0)} n^{-4} \right)^{1/6} \tag{Equation 2}
\]

where \( J(\lambda) \) is the overlap integral of the donor emission with the acceptor absorption spectrum (M\(^{-1}\) cm\(^{-1}\) nm\(^4\)), \( \kappa^2 \) accounts for the relative orientation of the donor and acceptor (usually assumed to be 2/3), \( \Phi_{FD(0)} \) is the donor fluorescence quantum yield in absence of transfer (0.6 for Alexa 488 attached to DNA), and \( n \) is the refractive index of the medium (n=1.33). For the given units the constant \( c_{FT} \) equals 8.79 \( \times \) 10\(^{-5} \) mol.

For MFD-plots the efficiency of energy transfer is calculated from a number of different parameters, including the fluorescence intensities of the donor or acceptor (\( F_D \) or \( F_A \)), donor lifetime (\( \tau_{D(A)} \)) in the presence of acceptor, and from the anisotropy (\( r \)) of either the donor or the acceptor fluorophore (29,60). We determine the efficiency of the energy transfer using the fluorescence intensity ratio between the donor and acceptor (\( F_D/F_A \)) (18). For analysis of these data the signal intensities (\( S_G \) and \( S_R \)) were corrected for background counts (typically 1.5-3.0 kHz for the green channels, \( B_G \) (donor), and 0.5-1.5 kHz for the red channels, \( B_R \) (acceptor)), spectral crosstalk, \( \alpha \) (0.058), and the ratio of the detection efficiencies, \( g \), between the green and red channels (\( g_G/g_R = 0.78 \)) (Equation 3).

\[
F_D = \frac{S_G - B_G}{g_G} = \frac{F_G}{g_G} \tag{Equation 3}
\]

\[
F_A = \frac{(S_R - B_R) - \alpha (S_G - B_G)}{g_R} = \frac{F_R}{g_R}
\]

For the specific dye pair used in this study, a small fraction of the recorded counts in the acceptor channel are due to direct excitation of the acceptor by the incoming laser light (let these counts be denoted as \( DE \)). In PDA an accurate determination of the FRET efficiency requires taking this effect into account. Thus equation 3 should be rewritten as follows:

\[
F_A = \frac{S_R - \left( \langle B_R \rangle + DE \right) - \alpha (S_G - \langle B_G \rangle)}{g_R} = \frac{F_R}{g_R} \tag{Equation 4}
\]
If direct excitation is not considered, we refer to the FRET efficiency $E$ as apparent FRET efficiency $E_a$. If the quantum yield of the acceptor, $\Phi_{FA}$, is taken into account (0.9 as measured for Alexa 594 attached to DNA), the efficiency of the energy transfer, $E$, can be calculated as:

$$E = \frac{F_A}{\Phi_{FA}} - \frac{F_A}{\Phi_{FA} + \frac{F_A}{\Phi_{FD(0)}}}$$  \hspace{1cm} (Equation 5)

It is important to note, that in MFD-plots the apparent FRET efficiency $\langle E_a \rangle$ is computed without correcting for direct acceptor excitation so that the apparent FRET efficiencies in MFD-plots are slightly higher than the FRET efficiencies $\langle E \rangle$ obtained by PDA, where the correction for direct excitation is implemented. The differences between $E_a$ and $E$ are mainly visible for small FRET efficiencies as measured in this work.

By combining Equation 1 and Equation 5 the analysis of the interdye distance can be performed by the help of the following equation:

$$R_{DA} = R_0 \left( \frac{\Phi_{FA}}{\Phi_{FD(0)}} \frac{F_D}{F_A} \right)^{\frac{1}{6}}$$  \hspace{1cm} (Equation 6)

For the data analysis by equation 1 and 5 we experimentally determined the Förster radius ($R_0 = 53.2 \text{ Å}$), for the Alexa 488-Alexa 594 pair.

The fluorescence lifetime is determined for each burst in two steps: (i) by generating a histogram of photon arrival times, and (ii) by fitting the histograms to a single exponential using a maximum likelihood estimator and iterative convolutions to account for the scatter contribution (61). The lifetime of the donor molecule in the absence of the acceptor ($\tau_{D(0)}$) was determined to be 3.8 ns. The efficiency of energy transfer is related to fluorescence lifetimes through (Equation 5):

$$E = 1 - \frac{\tau_{D(A)}}{\tau_{D(0)}}$$  \hspace{1cm} (Equation 7)
where \( \tau_{D(A)} \), \( \tau_{D(0)} \) is the lifetime of the donor in the presence and in absence of the acceptor respectively.

Another useful parameter calculated for each burst is the steady-state polarization anisotropy (62-63) of the donor, \( r_D \). This value is determined by using Equation 8:

\[
\frac{gS_{\|} - S_{\perp}}{(1-3l_1)gS_{\|} + (2-3l_2)S_{\perp}}
\]

(Equation 8)

The ratio of the detection efficiencies of the detection of the parallel and the perpendicular polarized photons is given as \( g \). The factors \( l_1 \) and \( l_2 \) account for polarization mixing because of the objective as described (63). The anisotropy is linked to the donor lifetime via Perrin’s equation:

\[
r_D = r_0 \left( \frac{\tau_{D(A)}}{\rho_D} + 1 \right)^{-1}
\]

(Equation 9)

In Equation 7, \( \rho_D \) is the rotational correlation time of D and \( r_0 \) is the fundamental anisotropy. Strictly speaking, the Perrin equation is valid only if the anisotropy decay is single-exponential, which is usually not the case. Still, the rotational correlation time in Equation 7 qualitatively reflects the rotational mobility of the dye, which is the collective result of the rotation of the dye molecule itself and the overall rotation of the biomolecule to which the dye is attached. In a 2-D plot of \( r_D \) vs \( \tau_{D(A)} \) the theoretical relation described in Perrin’s equation can be illustrated as a curve (Perrin curve) and an estimation of the mobility can be obtained.

2.3 Photon Distribution Analysis (PDA)

Photon distribution analysis (PDA) calculates the theoretical probability of recording a particular combination of red and green photons for a given FRET efficiency, and thus distance, can be derived from this analysis (19). PDA quantitatively describes the shapes of FRET distributions measured by MFD, including the effect of background and shot-noise in the distribution. Both the mean and width of a distribution are the functions of the mean FRET efficiency (31). With the PDA approach, a single parameter, which corresponds to the mean FRET efficiency, determines automatically the maximum width and asymmetry of the distribution that can be
assigned to shot-noise. Any additional broadening is attributed to the real interdye distance
distribution, which can reveal biologically relevant heterogeneities in an ensemble of
biomolecules (19).

2.3.1 Fitting

Each individual burst was divided into time windows of 2ms and histograms of the recovered
FRET efficiencies for each time window were created by using Equations 3, 4 and 5. By
inspection of the histograms for various time windows one could identify the presence of any
dynamic processes occurring in the ~ms time scale. In the current study both PDA analysis with
varying time windows and also FCS data revealed no dynamic behavior in the ~ms range.

General information about the fitting procedure of FRET efficiency distributions with PDA is
presented below, followed by specific details about the fits used in this study. The choice of the
form of the model function is crucial when fitting the data, since multiple FRET-efficiency states
are commonly observed in the experimental histograms. That is the fitting procedure proceeds in
conjunction with testing of alternative hypotheses regarding the model function (22,64). In the
procedure it is important that the number of free parameters used in each fit is minimized by
exploiting additional measurements (i.e. measurements of Donor only).

The experimental histograms in this study were fitted using in-house developed software. The
model functions in all cases were given in interdye distances, \( \langle R_{DA} \rangle_E \). A small percentage of
direct excitation of the acceptor (\( p_{DE} = 3.5\% \) as determined in (34) was also taken into account
during the fitting procedure).

For all measurements it was statistically justified to use a model function containing 3 states for
achieving a good fit (see SI Fig.1 and main text Fig 2-4). These populations accounted for Donor
only (\( D_{\text{only}} \), green line in the main text figures 2 and 4), one or two of the FRET-states (LF, MF,
HF) and in the measurements, for which no HF was present, impurities (dark brown line). In all
measurements there was a population of \( D_{\text{only}} \), which has very low \( \langle E \rangle \), accounts for around 20-
50\% of the total number of recorded bursts and it could be fitted by a single FRET apparent
distance (\( \langle R_{DA} \rangle_E = 110 \) Å as was determined by independent measurements and fixed for all fits).
For all measurements, a Gaussian distribution of interdye distances was used for each of the other two fitted states present in the model function (determined by mean value: \( \langle R_{DA} \rangle_E \) and halfwidth: HW). In measurements of \( G_A:T_D \) (main text Figure 2F), \( T_A:G_D +\text{MutS} \) (main text Figure 2F), \( G_A:C_D +\text{MutS} \) (main text Figure 2J) the majority of the population belongs to one FRET state (either LF or MF) and the third state accounts for a small fraction of impurities. The population due to impurities varied between 3-10 % and exhibited a broad FRET distribution with \( \langle R_{DA} \rangle_E \) close to the HF population as determined by independent measurement. Please note that this distribution was never populated to a significant extent and careful examination of “bursts” arising from the buffer indicated that this minor population is indeed arising from impurities in the preparation of the buffer. In the cases in which HF population was present the impurities could not be retrieved because they overlap with the strongly populated HF state. In principle only the shoulder of impurity populations is visible in the PDA histogram, because its peak is hidden in the presence of the HF population. Thus in the cases \( G_A:T_D +\text{MutS} \) (main text Figure 2G), \( G_A:G_D +\text{MutS} \) (main text Figure 2I) the fitting function, apart from \( D_{\text{only}} \), respectively included HF and LF in the former case and HF and MF in the latter. All the above comments could be summarized and referred to what we will note as:

**Fitmodel 1:** Initially only the Donor-only population was fixed and the other two states were left free to vary (3 free parameters per Gaussian distribution: mean distance \( \langle R_{DA} \rangle_E \), Half width HW, Amplitude (A)). For all measurements the fitting procedure was successful with low \( \chi^2 \)-values as it is shown in Supplementary Tables S1.1-1.3. Some concerns arose only for the cases of \( G_A:T_D-\text{MutS-ATP} \) and \( G_A:T_D-\text{MutS-ADP} \), since in these cases the fits were unstable regarding the low FRET population as illustrated by the large confidence intervals retrieved for these populations.

The results by using **Fitmodel 1** are presented in tables (see tables S1.1- S1.6). In tables S1.1-S1.2 the model function for each measurement is presented together with the corresponding \( \chi^2 \). In the table S1.3 one could evaluate the stability of the fits of some representative measurements by inspecting the 68% confidence intervals of all parameters. Moreover the identification of the individual FRET relevant states LF, MF and HF could be judged by inspecting tables S1.4- 1.6. One could observe small variations for the mean distance and HW of each state. The identified states LF, MF and HF are well discriminated as demonstrated by the values of the mean distances and the corresponding values of the FRET efficiencies (LF: \( \langle R_{DA} \rangle_E = 82.3 \pm 1.3 \) Å,
\[ \langle E \rangle = 0.070 \pm 0.007, \text{MF:} \langle R_{DA}\rangle_E = 73.2 \pm 3.1 \text{ Å}, \langle E \rangle = 0.17 \pm 0.04, \text{HF:} \langle R_{DA}\rangle_E = 63.1 \pm 0.8 \text{ Å}, \langle E \rangle = 0.27 \pm 0.01). \]

An argument could be made that FRET state 1 in the case of G_A:T_D-MutS-ADP is different than in the case of T_A:G_D-MutS (see Supplementary Table S1.1-2 and S1.5). These states are indeed different as proven by analysis of additional fluorescence parameters like donor anisotropy (see main text section “smMFD analysis of DNA-MutS complexes”), which is high in the T:G-MutS MF population but not in the G_A:T_D-MutS-ADP MF population (Figure 2 and Figure 5 in the main text and). Nevertheless, in order to describe PDA histograms with a minimal number of states following the parsimony principle, no additional states were included; thus the middle FRET population for G_A:T_D-MutS-ADP and T_A:G_D-MutS were considered to be identical for the purpose of data fitting. The HF population for G_A:T_D, G_A:T_D-MutS-ADP and G_A:T_D-MutS-ATP has \( \langle R_{DA}\rangle_E \) values of 62.3 Å, 64.2 Å and 63.2 Å. Since the differences in \( R_{\text{mean}} \) are small and all values of \( \langle R_{DA}\rangle_E \) lie within the confidence intervals of each other (Figure S1), we have chosen to consider these states as one. As an alternative, we preferred a fitmodel (which yielded equally good fits as shown in Supplement Table 1.9), where the relevant FRET states, were identified from the measurements in which they were predominant and only their relative fractions were left to vary.

This fitmodel which we will denote as **Fitmodel 2** could be described as follows:

**Fitmodel 2:** The FRET state LF (\( \langle R_{DA}\rangle_E = 82.7 \text{ Å}, \text{HW} = 7.1 \text{ Å} \)) was determined by taking the average values for \( \langle R_{DA}\rangle_E \) and HW of the distributions obtained in the measurements of free DNAs (G_A:T_D, G_A:G_D, G_A:C_D, T_A:G_D)). HF and MF were fixed to the values retrieved for the corresponding measurements of G_A:T_D+MutS and T_A:G_D+MutS (HF: \( \langle R_{DA}\rangle_E = 62.3 \text{ Å}, \text{HW} = 4.1 \text{ Å} \) and MF: \( \langle R_{DA}\rangle_E = 75 \text{ Å}, \text{HW} = 9.3 \text{ Å} \)). Only the relative fractions of these states were left to vary in each fit. The HF population was always consisted of a narrow distribution (HW ~ 4 Å) as a result in the cases of G_A:T_D, G_A:T_D-MutS-ADP and G_A:T_D-MutS-ATP has \( \langle R_{DA}\rangle_E \) values of 62.3 Å, 64.2 Å and 63.2 Å were used since when using the fixed value of HF the \( \chi^2 \) got significantly worse. Nevertheless, by inspecting table S.1.6 it is clear that all these high FRET distributions describe the same state (the variation in the mean distance value is only 0.8 Å).
The results of the fits with Fitmodel 2 are more robust (compare table S1.9 and table S1.3), yielding more restricted confidence intervals with comparable $\chi^2_r$-values as compared to the fits with Fitmodel 1. Thus, all results presented in the main text are from Fitmodel 2. All the results of fits with both Fitmodel 1 and 2 are presented in the supplement tables S1.1- S1.9, in order to illustrate that fixing the states does not alter qualitatively the results of the fits. Although 5 states are needed to describe all data for all different measurements, the number of free parameters per fit is limited as the D\textsubscript{only} peak is identified from independent measurements and because the fitting function for a single data set includes only 2 states, in addition to the fixed states for the D\textsubscript{only} population and the impurity population (6 free parameters per fit as shown in Supplementary Table S1.1 to S1.3 for Fitmodel 1 and 2-4 free parameters per experiment as shown in Supplementary Table S1.7-S1.9 for Fitmodel 2).

2.3.2 Error limits for multiple parameters

All free parameters of each fit were varied in a random manner around their optimal values. The $\chi^2_r$-value was calculated at 30000-55000 random points yielding more than 100 fits with $\chi^2_r$-values within $\pm (2/N_{\text{bins}})^{1/2}$ of the $\chi^2_r$-value of the optimal fit (the degrees of freedom of the fit were $N_{\text{bins}} = 55-59$) (65). Parameter intervals within these fits were assigned as confidence intervals, as presented in Supplementary Tables for some typical measurements.

2.4 Subensemble Fluorescence Correlation Spectroscopy (seFCS)

FCS is based on analysis of the fluctuations in fluorescence intensity (66-67). The origin of the fluctuations is not critical for the technique, thus FCS offers a unique tool for the direct study of phenomena like diffusion and photophysics and, via these phenomena, for the study of biologically relevant processes including macromolecular interactions (from the Stokes-Einstein relation it follows than the bigger a molecule the slower its diffusion in solution). Binding of the labeled oligonucleotides to MutS was investigated by FCS in this study, since unbound oligonucleotides diffuse faster than MutS-bound ones. seFCS curves were produced only for a specific subensemble by correlating signals from the green parallel and the green perpendicular detection channel of selected bursts, corresponding to individual FRET-states subpopulations as described in (61). Each correlation curve was fitted to a model function $G(t_{\text{lag}})$ with the lag (correlation) time $t_{\text{lag}}$ (see equation 10) including 2 bunching terms in the 1-300 µs scale.
accounting for the photophysics of the donor dye (triplet state kinetics, with a characteristic time of ~5 µs and a minor additional term around 250 µs which is assigned to dynamic fluorescence quenching). It should be noted that the data registration scheme utilized for smMFD enables use of the same data sets for correlation analysis as well as analysis of specific subpopulations of molecules. The diffusion times of the various FRET subpopulations were determined and compared in order to assess the free versus bound states of the molecules. The formula of the model function used is given below:

\[
G(t_{\text{lag}}) = 1 + \frac{1}{N} \cdot \frac{1}{1 + \frac{t_{\text{lag}}}{t_d}} \cdot \frac{1}{\sqrt{1 + \frac{t_{\text{lag}}}{\omega t_d^2}}} \cdot \left(1 - T_1 + T_1 \cdot e^{-t_{\text{lag}}/T_1} - T_2 + T_2 \cdot e^{-t_{\text{lag}}/T_2}\right) \quad (Equation \, 10)
\]

where \(N\) is the total number of molecules present in the detection volume, \(t_d\) is the apparent diffusion correlation time (slightly distorted by the burst cutting procedure), \(\omega\) is the geometrical parameter of the detection volume (66-67). \(T_1, T_2\) are the amplitudes and \(t_{T1}, t_{T2}\) the corresponding characteristic times of the two bunching terms.
Supplementary Table 1.1: Various mismatches. Fitting results for which the FRET populations are left free to vary (Fitmodel 1). LF: Blue background, MF: Orange background, HF: Magenta background, D\textsubscript{only}: Light green background, Impurities: Dark brown background.

| System          | Model function | \(\langle R_{DA}\rangle_E\) \(^1\) [Å] | HW \(^2\) [Å] | A \(^3\) | Rel. A \(^4\) [%] | \(\langle E\rangle \) \(^6\) | \(\langle R_{DA}\rangle_E\) \(^1\) [Å] | HW \(^2\) [Å] | A \(^3\) | Rel. A \(^4\) [%] | \(\langle E\rangle \) \(^6\) | \(\Delta\) \(^5\) A\text{\textsubscript{fixed}} |
|-----------------|----------------|--------------------------------------|--------------|--------|----------------|----------------|--------------------------------------|--------------|--------|----------------|----------------|------------------|
| GA:T\textsubscript{D} |                | 82.2                                 | 7.0          | 67.7   | 96.9          | 0.068          | 59.5                                 | 5.7          | 2.2    | 3.1            | 0.34           | 30.1             | 1.0              |
| GA:T\textsubscript{D}-MutS |        | 79.5                                 | 13.3         | 31.8   | 46.2          | 0.082          | 62.30                                | 4.1          | 37.1   | 53.8          | 0.28           | 31.1             | 1.2              |
| T\textsubscript{A}:G\textsubscript{D} |        | 82.3                                 | 6.6          | 42.6   | 93.0          | 0.068          | 60.9                                 | 9.0          | 3.2    | 7.0            | 0.31           | 54.2             | 0.75             |
| T\textsubscript{A}:G\textsubscript{D}-MutS |        | 75                                   | 9.3          | 68.8   | 98.4          | 0.11           | 49.9                                 | 1.8          | 1.1    | 1.6            | 0.60           | 30.1             | 1.5              |
| GA:G\textsubscript{D} |                | 83.5                                 | 7.6          | 39.3   | 98.5          | 0.063          | 56                                   | 7.2          | 0.6    | 1.5            | 0.42           | 60.1             | 1.1              |
| GA:G\textsubscript{D}-MutS |        | 75.1                                 | 11.4         | 43.1   | 67.3          | 0.11           | 63.2                                 | 3.3          | 20.9   | 32.7          | 0.26           | 36               | 1.1              |
| GA:C\textsubscript{D} |                | 82.7                                 | 7.2          | 59.1   | 91.3          | 0.066          | 74.3                                 | 18.3         | 5.6    | 8.7            | 0.12           | 35.3             | 0.94             |
| GA:C\textsubscript{D}-MutS |        | 83.3                                 | 7.9          | 65.7   | 94.0          | 0.064          | 63.7                                 | 9.4          | 4.2    | 6.0            | 0.25           | 30.1             | 0.84             |

\(^1\)\(\langle R_{DA}\rangle_E\) corresponds to the FRET-averaged distance between the dyes, \(^2\)HW to the half width of the distribution (see Supplementary Section 2.3.1).\(^3\)A refers to the amplitude of a given population as given by the fit, \(^4\)rel. A refers to the relative amplitude of the population in relation only to the FRET populations (Donor-only excluded). \(^5\)A\text{\textsubscript{fixed}} refers to the Donor-only population, considered fixed at 110Å, \(^6\)\(\langle E\rangle\) corresponds to the FRET efficiency value for Förster radius of 53.2 Å and an interdye distance \(\langle R_{DA}\rangle_E\).
Supplementary Table 1.2: Nucleotide Effect. Results of fits for which the FRET populations are left free to vary (Fitmodel 1). LF: Blue background, MF: Orange background, HF: Magenta background, D_only: Light green background, Impurities: Dark brown background.

| System                  | Model function | State 1                  | State 2                  | D_only | \(\chi^2\) |
|-------------------------|----------------|--------------------------|--------------------------|--------|------------|
|                         |                | \(\langle R_{DA}\rangle_E\) | HW | A | Rel. A | \langle E\rangle | HW | A | Rel. A | \langle E\rangle | \(A_{\text{fixed}}\) |
| **G_A:T_D**             |                | 82.2                     | 7.0 | 67.7 | 96.9 | 0.068 | 59.5 | 5.7 | 2.2 | 3.1 | 0.34 | 30.1 | 1 |
| **G_A:T_D-MutS**        |                | 79.5                     | 13.3 | 31.8 | 46.2 | 0.082 | 62.3 | 4.1 | 37.1 | 53.8 | 0.28 | 31.1 | 1.23 |
| **G_A:T_D-MutS-AMP-PNP**|                | 82.4                     | 6.1 | 69.1 | 95.8 | 0.068 | 60.7 | 4.5 | 3.0 | 4.2 | 0.31 | 27.9 | 1.51 |
| **G_A:T_D-MutS-ATP**    |                | 80.2                     | 12.1 | 25.3 | 33.9 | 0.079 | 64.1 | 4.2 | 49.4 | 66.1 | 0.25 | 25.3 | 0.79 |
| **G_A:T_D-MutS-ADP**    |                | 69.6                     | 9.2 | 32.1 | 43.5 | 0.17  | 62.8 | 2.5 | 41.7 | 56.5 | 0.27 | 26.2 | 0.8 |

1\(\langle R_{DA}\rangle_E\) corresponds to the FRET-averaged distance between the dyes, 2HW to the half width of the distribution (see Supplementary Section 2.3.1). 3A refers to the amplitude of a given population as given by the fit. 4rel. A refers to the relative amplitude of the population in relation only to the FRET populations (D_only excluded). 5\(A_{\text{fixed}}\) refers to the donor-only population, considered fixed at 110 Å, 6\(\langle E\rangle\) corresponds to the FRET efficiency value for Förster radius of 53.2 Å and an interdye distance \(\langle R_{DA}\rangle_E\).
**Supplementary Table 1.3: Nucleotide Effect.** Results of fits for which the FRET populations are left free to vary (Fitmodel 1). In black the 95% confidence intervals are given for each parameter. In the parenthesis in red the value of each parameter corresponding to the best fit is given for direct comparison. The columns of the mean distance and the corresponding mean FRET efficiency are shown in bold. The confidence intervals were calculated as described in Supplementary Section 2.3.2. Similar confidence intervals were obtained for the measurements not described in this table. LF: Blue background, MF: Orange background, HF: Magenta background, D only: Light green background, Impurities: Dark brown background.

| System                  | Model Function (Only Gaussian Distributions) | State 1 | State 2 | \(\chi^2_r\) |
|-------------------------|---------------------------------------------|---------|---------|--------------|
|                         | \(\langle R_{DA}\rangle_E\)^1 | HW^2 | Rel. A^3 | \(\langle E\rangle^4\) | \(\langle R_{DA}\rangle_E\)^1 | HW^2 | Rel. A^3 | \(\langle E\rangle^4\) |
|                         | [Å] | [Å] | [%] |         | [Å] | [Å] | [%] |         |
| **GA:TD**               | 81.1-83.5 | 5.8-7.8 | 92.5-98.1 | 0.06-0.07 | 55.0-69.3 | 4.8-8.9 | 1.9-7.5 | 0.17-0.45 | 0.91-1.29 |
|                         | (82.2) | (6.8) | (95.7) | (0.07) | (61.7) | (6.9) | (4.3) | (0.29) |
| **GA:TD-MutS**          | 75.7-85.4 | 9.8-16.2 | 35.5-54.3 | 0.06-0.10 | 61.4-63.3 | 3.3-4.9 | 45.7-64.5 | 0.26-0.3 | 1.13-1.52 |
|                         | (79.5) | (13.2) | (46.1) | (0.08) | (62.3) | (4.1) | (53.9) | (0.28) |
| **TA:GD-MutS**          | 74.2-76.0 | 8.5-10.6 | 96.9-100 | 0.11-0.12 | 41.3-59.8 | 0.4-3.7 | 0.0-3.1 | 0.33-0.82 | 1.36-1.75 |
|                         | (75.1) | (9.3) | (98.5) | (0.11) | (50.0) | (2.1) | (1.5) | (0.60) |
| **GA:TD-MutS-ADP**      | 67.0-73.7 | 7.2-11.6 | 29.3-56.4 | 0.12-0.20 | 62-63.5 | 2.1-2.8 | 43.6-70.7 | 0.26-0.29 | 0.74-1.13 |
|                         | (69.6) | (9.2) | (43.5) | (0.17) | (62.8) | (2.5) | (56.5) | (0.27) |
| **GA:TD-MutS-ATP**      | 74.6-88.1 | 8.4-15.7 | 19.8-50.1 | 0.05-0.12 | 63.1-65.2 | 3.4-5.1 | 49.9-80.2 | 0.23-0.26 | 0.72-1.11 |

^1\(\langle R_{DA}\rangle_E\) corresponds to the FRET-averaged distance between the dyes, ^2HW to the half width of the distribution (see Supplementary Section 2.3.1), ^3Rel. A refers to the relative amplitude of the population in relation only to the FRET populations (D only excluded). ^4\(\langle E\rangle\) corresponds to the FRET efficiency value for Förster radius of 53.2 Å and an interdye distance \(\langle R_{DA}\rangle_E\).
Supplementary Table 1.4: Identification of LF state from Fitmodel 1.

| System          | Low FRET state (LF) | 
|-----------------|---------------------|
|                 | $\langle E \rangle$ | $\langle R_{D,0}(E) \rangle$ [Å] |
| G:C             | 0.066               | 82.7 |
| G:C+MutS        | 0.064               | 83.3 |
| G:T             | 0.068               | 82.2 |
| G:T+MutS        | 0.082               | 79.5 |
| T:G             | 0.068               | 82.3 |
| G:G             | 0.063               | 83.5 |
| G:T+MutS+AMPnP  | 0.068               | 82.4 |
| G:T+MutS+ATP    | 0.079               | 80.2 |

| Mean value      | 0.070               | 82.3 |
| Standard deviation | 0.007               | 1.3 |

Supplementary Table 1.5: Identification of MF state from Fitmodel 1.

| System          | Middle FRET state (MF) | 
|-----------------|------------------------|
|                 | $\langle E \rangle$ | $\langle R_{D,0}(E) \rangle$ [Å] |
| T:G+MutS        | 0.11                  | 75 |
| G:T+MutS+ADP    | 0.17                  | 69.6 |
| G:G+MutS        | 0.1                   | 75.1 |

| Mean value      | 0.127                | 73.2 |
| Standard deviation | 0.038               | 3.1 |

Supplementary Table 1.6 Identification of HF state from Fitmodel 1.

| System          | High FRET state (HF) | 
|-----------------|----------------------|
|                 | $\langle E \rangle$ | $\langle R_{D,0}(E) \rangle$ [Å] |
| G:T+MutS        | 0.28                 | 62.3 |
| G:T+MutS+ATP    | 0.25                 | 64.1 |
| G:T+MutS+ADP    | 0.27                 | 62.8 |
| G:G+MutS        | 0.26                 | 63.2 |

| Mean value      | 0.265                | 63.1 |
| Standard deviation | 0.013               | 0.8 |
**Supplementary Table 1.7: Different mismatches.** Results of fits for which the FRET populations were fixed (Fitmodel 2) as described in the text (Supplementary Section 2.3.1). LF: Blue background, MF: Orange background, HF: Magenta background, D_only: Light green background, Impurities: Dark brown background.

| System          | Model function | State 1 | State 2 | D_only |
|-----------------|----------------|---------|---------|--------|
|                 |                | $\langle R_{DA}\rangle_E$ | HW$^2$ | A$^3$ | Rel. A$^4$ | $\langle E \rangle^6$ | $\langle R_{DA}\rangle_E$ | HW$^2$ | A$^3$ | Rel. A$^4$ | $\langle E \rangle^6$ | $\chi^2_\nu$ |
| $G_A:T_D$       |                | 82.7    | 7.1     | 66.3   | 94.0   | 0.066 | 65.4 | 5.7 | 4.2 | 6.0 | **0.23** | 29.5 | 0.99 |
| $G_A:T_D$-MutS  |                | 82.7    | 7.1     | 23.3   | 34.6   | 0.066 | 62.3 | 4.1 | 44  | 65.4 | **0.28** | 32.7 | 1.75 |
| $T_A:G_D$       |                | 82.7    | 7.1     | 43.4   | 93.5   | 0.066 | 60.5 | 9.0 | 3   | 6.5 | **0.32** | 53.6 | 0.76 |
| $T_A:G_D$-MutS  |                | 75      | 9.3     | 68.9   | 98.9   | 0.11  | 49.6 | 1.4 | 0.8 | 1.1 | **0.60** | 30.3 | 1.46 |
| $G_A:G_D$       |                | 82.7    | 7.1     | 38.1   | 97.4   | 0.066 | 58.3 | 9.6 | 1   | 2.6 | **0.37** | 60.9 | 1.04 |
| $G_A:G_D$-MutS  |                | 75      | 9.6     | 41.3   | 65.8   | 0.11  | 62.3 | 4.1 | 21.5| 34.2 | **0.28** | 37.2 | 1.38 |
| $G_A:C_D$       |                | 82.7    | 7.1     | 58.2   | 90.0   | 0.066 | 76.5 | 18.7| 6.5 | 10.0 | **0.10** | 35.3 | 0.91 |
| $G_A:C_D$-MutS  |                | 82.7    | 7.1     | 65.02  | 94.8   | 0.066 | 62.3 | 4.1 | 3.6 | 5.2 | **0.28** | 31.4 | 1.16 |

$^1\langle R_{DA}\rangle_E$ corresponds to the FRET-averaged distance between the dyes, $^2$HW to the half width of the distribution (see Supplementary Section 2.3.1). $^3$A refers to the amplitude of a given population as given by the fit, $^4$rel. A refers to the relative amplitude of the population in relation only to the FRET populations (D_only excluded). $^5A_{\text{fixed}}$ refers to the donor-only population, considered fixed at 110Å, $^6\langle E \rangle$ corresponds to the FRET efficiency value for Förster radius of 53.2 Å and an interdye distance $\langle R_{DA}\rangle_E$. 
**Supplementary Table 1.8: Nucleotide Effect.** Results of fits for which the FRET populations were fixed (Fitmodel 2) as described in the text (Supplementary section 2.3.1). LF: Blue background, MF: Orange background, HF: Magenta background, \( D_{\text{only}} \): Light green background, Impurities: Dark brown background.

| System | Model Function | \( \chi^2 \) | \( \chi^2 \) |
|--------|----------------|----------------|----------------|
| \( \text{G}_A: \text{T} \) | | | |
| \( \langle R_{\text{DA}} \rangle_E \) | HW\(^1\) | \( \langle E \rangle \) \(^6\) | HW\(^1\) | \( \langle E \rangle \) \(^6\) | 27.4 | 1.38 |
| \( \text{A} \) | [Å] | [Å] | | [Å] | [Å] | |
| \( \text{rel. A} \) | [%] | [%] | | [%] | [%] | |
| \( \text{State 1} \) | \( \text{State 2} \) | \( \text{D}_{\text{only}} \) | \( \text{D}_{\text{only}} \) |
| \( \text{GA:TD} \) | 82.7 | 7.1 | 71.3 | 98.2 | 0.07 | 58.1 | 5.6 | 1.3 | 1.8 | 0.37 | 27.4 | 1.38 |
| \( \text{GA:TD} - \text{MutS} \) | 82.7 | 7.1 | 16.6 | 24.7 | 0.07 | 62.3 | 4.1 | 51 | 75.3 | 0.28 | 32.7 | 1.51 |
| \( \text{GA:TD} - \text{MutS-AMP-PNP} \) | 82.7 | 7.1 | 70.9 | 97.3 | 0.07 | 58.5 | 3.2 | 2 | 2.7 | 0.36 | 27.1 | 1.56 |
| \( \text{GA:TD} - \text{MutS-ADP} \) | 75 | 9.6 | 18.2 | 24.4 | 0.11 | 63.2 | 3.5 | 56 | 75.6 | 0.26 | 25.5 | 0.92 |
| \( \text{GA:TD} - \text{MutS-ATP} \) | 82.7 | 7.1 | 18.4 | 25.0 | 0.07 | 64.2 | 4.5 | 55 | 75.0 | 0.25 | 26.4 | 0.91 |

\(^1\langle R_{\text{DA}} \rangle_E \) corresponds to the FRET-averaged distance between the dyes, \(^2\text{HW}\) to the half width of the distribution (see Supplementary Section 2.3.1).\(^3\text{A}\) refers to the amplitude of a given population as given by the fit, \(^4\text{rel. A}\) refers to the relative amplitude of the population in relation only to the FRET populations (\( D_{\text{only}} \) excluded). \(^5\text{A}_{\text{fixed}}\) refers to the donor-only population, considered fixed at 110Å, \(^6\langle E \rangle \) corresponds to the FRET efficiency value for Förster radius of 53.2 Å and an interdye distance \( \langle R_{\text{DA}} \rangle_E \).
**Supplementary Table 1.9: Nucleotide effect.** The 95% confidence intervals of fits for which the FRET populations were fixed (Fitmodel 2) as described in the text (Supplementary Section 2.3.1). The parameters for which confidence intervals are not given were fixed during the fit. The $\chi^2$-values corresponding to good fits are given for the case of the fits with Fitmodel 1 and 2 for direct comparison. The confidence intervals were calculated as described in Supplementary Section 2.3.2. Similar confidence intervals were obtained for the measurements not described in this table. LF: Blue background, MF: Orange background, HF: Magenta background, Donly: Light green background, Impurities: Dark brown background.

| System            | Model Function | $\chi^2$ (Fitmodel 2) | $\chi^2$ (Fitmodel 1) |
|-------------------|----------------|-----------------------|-----------------------|
|                   | G_A:T_D        |                       |                       |
|                   |                | 82.7                  | 0.07                  |
|                   |                | 7.1                   |                       |
|                   |                | 89.2-97.4             |                       |
|                   |                | 57.0-75.1             |                       |
|                   |                | 6.2-11.5              |                       |
|                   |                | 2.6-11.0              |                       |
|                   |                | 0.11-0.40             |                       |
|                   |                | 0.94-1.32             |                       |
|                   |                | 0.91-1.29             |                       |
|                   | G_A:T_D-MutS   |                       |                       |
|                   |                | 82.7                  | 0.07                  |
|                   |                | 7.1                   |                       |
|                   |                | 30.5-38.8             |                       |
|                   |                | 62.3                  |                       |
|                   |                | 4.1                   |                       |
|                   |                | 61.2-69.5             |                       |
|                   |                | 0.28                  |                       |
|                   |                | 1.73-2.1              |                       |
|                   |                | 1.13-1.52             |                       |
|                   | T_A:G_D-MutS   |                       |                       |
|                   |                | 75                    | 0.11                  |
|                   |                | 9.3                   |                       |
|                   |                | 97.3-99.9             |                       |
|                   |                | 25.0-55.5             |                       |
|                   |                | 0.0-2.8               |                       |
|                   |                | 0.1-2.7               |                       |
|                   |                | 0.44-1.00             |                       |
|                   |                | 1.37-1.74             |                       |
|                   |                | 1.36-1.75             |                       |
|                   | G_A:T_D-MutS-ADP |                       |                       |
|                   |                | 75                    | 0.11                  |
|                   |                | 9.3                   |                       |
|                   |                | 17.3-32.9             |                       |
|                   |                | 63.0-63.9             |                       |
|                   |                | 1.6-4.7               |                       |
|                   |                | 67.1-82.7             |                       |
|                   |                | 0.26-0.45             |                       |
|                   |                | 0.99-1.37             |                       |
|                   |                | 0.74-1.13             |                       |
|                   | G_A:T_D-MutS-ATP |                       |                       |
|                   |                | 82.7                  | 0.07                  |
|                   |                | 7.1                   |                       |
|                   |                | 18.9-30.4             |                       |
|                   |                | 64.0-65.0             |                       |
|                   |                | 3.8-5.2               |                       |
|                   |                | 69.6-81.1             |                       |
|                   |                | 0.23-0.26             |                       |
|                   |                | 0.89-1.11             |                       |
|                   |                | 0.72-1.11             |                       |

$\langle \langle R_{DA} \rangle \rangle_E$ corresponds to the FRET-averaged distance between the dyes, $^2HW$ to the half width of the distribution (see Supplementary Section 2.3.1). $^3$rel. A refers to the relative amplitude of the population in relation only to the FRET populations (Donly excluded). $^4\langle E \rangle$ corresponds to the FRET efficiency value for Förster radius of 53.2 Å and an interdye distance $\langle R_{DA} \rangle_E$.  
Supplementary Figure 1: 2-state fits vs. 3-state fits for the case of G\textsubscript{A}:T\textsubscript{D}-MutS-ADP. The F-test returns that the 3-state model is justified with 100% probability. For the rest of measurements the need for using a 3-state model instead of a 2-state model is even more pronounced.

Supplementary Figure 2: Fluorescence correlation spectroscopy analysis of the smMFD data was used to determine and compare the diffusion times of the labeled DNA through the observation volume of each FRET species in the absence or presence of MutS. A: Typical correlation curves for the free and bound DNA oligonucleotides used in the single molecule measurements are shown in black and red correspondingly. B: Increased apparent diffusion correlation times in relation to the free DNA are indicative for the formation of a DNA-MutS complex. The color coding is the same as in the main text (Green: D-only, Blue: LF, Orange: MF, Magenta: HF).
Supplementary Figure 3: Sequence of unlabeled duplex DNA used in the competition experiments shown in Fig. 6 in the main text. The mismatched bases are shown in white letters on black background. DNA-MutS contacts are indicated for the case that MutS stacks only to the mismatched T. Contacts made by subunit A of MutS are shown in blue, contacts of subunit B in orange. Note that in case of T:G-6-MutS interaction less contacts are made because the mismatch is too close to the DNA end.

References

58. Eggeling, C., Berger, S., Brand, L., Fries, J.R., Schaffer, J., Volkmer, A. and Seidel, C.A. (2001) Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *J Biotechnol*, 86, 163-180.

59. Eggeling, C., Widengren, J., Brand, L., Schaffer, J., Felekyan, S. and Seidel, C.A. (2006) Analysis of photobleaching in single-molecule multicolor excitation and Forster resonance energy transfer measurements. *J Phys Chem A*, 110, 2979-2995.

60. Clegg, R.M. (1992) Fluorescence resonance energy transfer and nucleic acids. *Methods in Enzymol*, 211, 353-388.

61. Eggeling, C., Schaffer, J., Seidel, C.A.M., Korte, J., Brehm, G., Schneider, S. and Schrof, W. (2001) Homogeneity, Transport, and Signal Properties of Single Ag Particles Studied by Single-Molecule Surface-Enhanced Resonance Raman Scattering. *J Phys Chem A*, 105, 3673-3679.

62. Lakowicz, J.R. (2006) *Principles of Fluorescence Spectroscopy*. Third ed. Springer, New York.

63. Koshioka, M., Sasaki, K. and Masuhara, H. (1995) Time-Dependent Fluorescence Depolarization Analysis in 3-Dimensional Microspectroscopy. *Applied Spectroscopy*, 49, 224-228.
64. Torella, J.P., Holden, S.J., Santoso, Y., Hohlbein, J. and Kapanidis, A.N. (2011) Identifying molecular dynamics in single-molecule FRET experiments with burst variance analysis. Biophys J, 100, 1568-1577.

65. Philip, B.R. and Robinson, D.K. (2002) Data Reduction and Error Analysis for the Physical Sciences. McGraw Hill.

66. Ehrenberg, M. and Rigler, R. (1974) Rotational brownian motion and fluorescence intensify fluctuations. Chem Phys, 4, 390-401.

67. Magde, D., Elson, E. and Webb, W.W. (1972) Thermodynamic Fluctuations in a Reacting System Measurement by Fluorescence Correlation Spectroscopy. Phys Rev Lett, 29, 705.