Supplementary Information for:

Single-molecule imaging reveals target search mechanisms during DNA mismatch repair

Jason Gorman[2,*], Feng Wang[3‡], Sy Redding[4‡], Aaron J. Plys[5,†], Teresa Fazio[6], Shalom Wind[6], Eric Alani[5], and Eric C. Greene[1,3]

[1]Howard Hughes Medical Institute, and the Departments of [2]Biological Sciences, [3]Biochemistry and Molecular Biophysics, and [4]Chemistry, Columbia University, New York, NY, 10032; [5]Department of Molecular Biology and Genetics, Cornell University, 459 Biotechnology Building, Ithaca, NY 14853-2703; [6]Department of Applied Physics and Applied Mathematics, Center for Electron Transport in Molecular Nanostructures, NanoMedicine Center for Mechanical Biology, Columbia University 1020 Schapiro CEPSR, 530 West 120th St., New York, NY 10027

**Present Address: Vaccine Research Center, National Institutes of Health, Bethesda, MD, 20892*
†Present Address: Department of Biological Sciences, University of Cyprus, Nicosia, Cyprus
‡Equal Contribution.

To whom correspondence should be addressed: ecg2108@columbia.edu

Table of Contents

Materials and Methods

1. Protein purification and labeling. MutSα and MutLα were purified and labeled as described (1, 2). MutSα labeling was performed at a 6:1 QD:Protein ratio (300 nm Qdot : 50 nM protein) in PBS containing 0.2 mg ml^{-1} BSA and incubated for 20 minutes at 4°C. The protein-QD conjugates were then purified to remove unconjugated QDs. For this, biotinylated λ-DNA (300 pM) was incubated with streptavidin magnetic beads (5 mg; Roche) for 20-min at 20°C. The MutSα-QD conjugation reaction was added to the beads, the PBS solution was diluted to 1/5th concentration with 10 mM Tris (pH 7.8) solution, and the reaction was incubated for 10-min at 4°C. Beads were washed twice with 10 mM Tris (pH 7.8), 20 mM NaCl, 1 mM MgCl, 1 mM DTT, and 0.2 mg ml^{-1} BSA. QD-MutSα was eluted with 10 mM Tris (pH 7.8), 300 mM NaCl, 1 mM MgCl, 1 mM DTT, and 0.2 mg ml^{-1} BSA.

2. DNA substrates and cloning. To create λ-DNA with 3 tandem G/T mismatches a 151 bp DNA fragment containing unique restriction and nickase sites was ligated between the Nhe1 and XhoI sites (Fig. S1). Insert-containing DNA was packaged using MaxPlax λ packing extracts (Epicenter), according to the manufacturer's instructions. Phage stocks were prepared by standard plate lysis, and used to infect 1 ml of *E. coli* LE392MP cells (OD 0.1) at 37°C for 20 minutes. Infected cells were used to inoculate a 200 ml liquid culture in LB and 10 mM $MgSO_4$, which was grown overnight at 39°C. 10 ml of chloroform was added and the culture was shaken for 10 minutes. The lysed culture was incubated with Dnase I and RNase (1 μg ml^{-1} each) at 20°C for 1 hour. SDS (0.5%), EDTA (50 mM) and proteinase K (5 mg) were added to the lysed culture, and incubated at 20°C for 1 hour, followed by phenol chloroform extraction and isopropanol precipitation. Purified DNA was resuspended in TE, and end-labeled with oligonucleotides, as described (2). To make mismatches, the end-labeled DNA was treated with the Nt.BspQI (NEB), mixed with a 1000-fold molar excess of an oligonucleotide complementary to the region encompassed by the nickase sites, and then heated and cooled. Successful insertion was assessed by comparing restriction digests with either NcoI or SwaI, and alkaline gel electrophoresis verified the nicks were sealed by T4 DNA ligase.

3. Single molecule reaction conditions. Unless otherwise stated, reactions were performed as described (1, 2), with the exception that all buffers contained either 100 or 150 mM NaCl, unless otherwise specified, and the DNA substrates for both single- and double-tethered assays exhibited a mean extended contour length of ~ 0.75. In brief, buffers contained 20 mM Tris [pH 7.8], 1 mM $MgCl_2$, 1 mM DTT, and 4 mg ml^{-1} BSA, along with the indicated concentrations of NaCl. Unless otherwise stated, standard reaction conditions for looking at lesion binding all contained 1 mM ADP in the buffers. In the nucleotide chase experiments, the 1 mM ADP was replaced by injecting 1 mM ATP or 1 mM ATPγS, as specified. Unless otherwise, indicated YOYO1 was omitted. Please note that all reported results, and conclusions derived from these results, are based upon at least three independent experimental measurements.

4. Binding site distribution measurements on single-tethered DNA curtains. Binding distributions of MutSα and MutSα/MutLα complexes were made on single-tethered DNA curtains. Data was divided into bins based on the variation of QDs attached to DNA curtains at single fixed positions (3). The MutLα only distributions were obtained through the same analysis procedure, but used double-tethered DNA curtains (with no buffer flow), because MutLα diffuses rapidly along DNA and is quickly pushed off of single-tethered DNA curtains if flow is applied (2).

Because MutLα does not remain preferentially bound to any positions on the DNA (2), the distribution histogram of MutLα represents the instantaneous positions for all molecules in the observed population and the flat distribution reflects the absence of preferred binding sites. Sampling error was determined by the Bootstrap method (4), and the 70% confidence intervals are presented.

5. MutSα and MutLα target search experiments. These experiments were conducted with double-tethered curtains in 40 mM Tris (pH 7.8), 1 mM DTT, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM ADP, and 0.2 mg ml^{-1} BSA. For mismatch search experiments, QD-MutSα (1-5 nM) was injected at a flow rate of 5-20 μl/min, and flow was terminated upon visual confirmation that the proteins had begun entering the sample chamber. For the MutSα-mismatch search experiments, MutSα was pre-bound to the mismatch in buffer containing 1 mM ADP, and free proteins were flushed from the sample chamber. QD-MutLα (5-20 nM) was injected at a flow rate of 5-20 μl min^{-1}, and flow was terminated upon visual confirmation that the proteins had entered the sample chamber. A protein was categorized as having undergone a 1D search only if there were at least two frames at the beginning of the diffusion trajectory that were at least three standard deviations away from the location of the mismatch. If the proteins initially appeared within this resolution limit, then they were categorized has having undergone an apparent 3D binding event.

6. Mismatch-bound MutSα and MutSα/MutLα lifetime measurements on single-tethered DNA curtains. QD-tagged proteins were bound to mismatch-bearing DNA in single-tethered curtains in buffer containing 20 mM Tris [pH 7.8], 50 mM NaCl, 1 mM ADP, 1 mM $MgCl_2$, 1 mM DTT, and 4 mg ml^{-1} BSA. The NaCl concentration was raised to 150 mM and the images collected at defined intervals and the total number of proteins remaining bound to the lesions was plotted as a function of time. Resulting data were fit to single exponential curves.

7. Nucleotide chase experiments on double-tethered DNA curtains. ATP chase experiments for were performed with double-tethered DNA curtains, so buffer flow could be terminated after ATP injection, or maintained at a very low constant rate (5-20 μl/min) such that the diffusive properties of the DNA-bound proteins were not perturbed. The delay time of the injection system was pre-calibrated by using the microscope to monitor the background fluorescence signal following injection of fluorescein. QD-tagged proteins (MutSα or the MutSα/MutLα complex, as indicated) were first bound to mismatch-bearing DNA molecules in buffer containing 20 mM Tris [pH 7.8], 150 mM NaCl, 1 mM ADP, 1 mM $MgCl_2$, 1 mM DTT, and 4 mg ml^{-1} BSA, and the reactions were chased with the same buffer but with the 1 mM ADP replaced with 1 mM ATP (or 1 mM ATPγS, as indicated). Videos were continually recorded at 5- or 10-Hz, and the data manually segregated into populations that either remained stationary, directly dissociated from the DNA, or began diffusing along the DNA.

8. Protein tracking and diffusion coefficients. Diffusion coefficients represent the mean ± standard deviation of ≥25 particle tracking measurements and were calculated from MSD plots as described (1, 2). All diffusion coefficients were based on measurements of protein complexes that exhibited QD blinking (see below); the reason these measurements are confined to blinking QDs is to help ensure that the reported diffusion coefficients reflect a homogeneous population of molecules all with the same hydrodynamic radii, and minimize variance associated with the

reported values is due to heterogeneity in the oligomeric states of the complexes being measured (1, 2). The spatial resolution of our tracking data is limited by Brownian fluctuations of the DNA. To determine the scale of this noise, QDs were attached to the DNA through a digoxignein linker. The λ-DNA for these curtains was engineered to include an oligonucleotide insertion containing digoxigenin and the curtains were incubated with QDs labeled with anti-digoxigenin antibodies, and the standard deviation of the QDs was calculated for 120 molecules over 4 different flowcells. The average of the 120 measured standard deviations, representing the average longitudinal fluctuations of the DNA molecules, was 30-nanometers (~120-bp). Note that observed diffusion coefficients displayed a log-normal distribution, as expected given that the energy landscape on the DNA experienced by the diffusing proteins is normally distributed (1, 2). A Student's t-test can only evaluate normal distributions, therefore when comparing two different diffusion coefficients the p-values were obtained from the natural logs (ln) of the corresponding diffusion coefficients, as described (2).

Individual QDs blink, and this well-known phenomenon enables one to distinguish single vs. multiple QDs (2, 5-7). In our experiments a non-blinking QD signal could arise from either QD aggregation, or protein aggregation/oligomerization; note that we cannot accurately determine the numbers of proteins that might be present in these larger complexes because as many as 50% of the QDs can be dark (*i.e.*: non-fluorescent), and the emission intensity of individual QDs can also span a broad range (6). From our experiments, ≥95% of all observed molecules of QD-MutSα (either diffusing along the DNA or bound to the mismatches) exhibited QD blinking, as would be expected from single QD-tagged proteins, and as previously reported, ~74% of all observed QD-MutLα exhibited blinking before binding to MutSα.(2) Of the remaining ~26% of QD-MutLα that did not blink, these molecules exhibited signal intensities that could be consistent with larger oligomers of MutLα, and many of these (52%) also diffused on DNA, as previously reported (2). For the MutSα only binding distribution measurements, we only measured the locations of blinking QDs, which again reflected ≥95% of the total population of molecules. For the MutLα only binding distribution measurements, we only measured the locations of blinking QDs, which again reflected ~74% of the total population. For the MutLα binding distribution measurements made in the presence of MutSα (either QD-tagged MutSα or untagged MutSα), we did not segregate the MutLα data based on QD blinking behavior, and the resulting distribution histograms were representative of the entire population of observed proteins. The reason we did not segregate this MutLα binding site distribution data into blinking and nonblinking populations is because once the first MutSα/MutLα complex has formed at a lesion, other upstream MutLα molecules are pushed into this stationary complex due to the force exerted on the proteins by the flowing buffer (see also Gorman *et al.*, 2010),(2) and these incoming proteins tend to stack up at the lesions because they cannot be pushed passed the lesion bound complex. We do not interpret this MutLα "stacking behavior" seen in the single-tethered curtain assays as oligomerization of MutLα at lesion-bound MutSα, because when the same experiments are conducted on double-tethered curtains (in the absence of buffer flow) additional diffusing molecules of MutLα do not appear to oligomerize after formation of the initial lesion-bound MutSα/MutLα complex, and under these conditions 90% of all observed lesion-bound complexes display blinking of the QD-MutLα (see below). For the MutSα target search, MutSα only nucleotide chase experiments, and spontaneous mismatch escape experiments, ≥95% of all QD-MutSα molecules displayed blinking, and only these blinking molecules were used for analysis. For the MutLα target search experiments, ~74% of all QD-MutLα scanning the DNA

also exhibited blinking behavior, and we confined our analysis to these proteins, although as previously reported the nonblinking fraction of QD-MutLα also diffused along the DNA by 1D diffusion, and we saw no obvious differences in the behavior of blinking versus nonblinking QD-MutLα complexes. For the MutSα/MutLα complex ATP chase experiments, 100% of the QD-MutSα (N=39/39) and 90% of the QD-MutLα (N=35/39) within the mismatch-bound MutSα/MutLα complexes exhibited blinking behavior consistent with single proteins. The remaining 10% of the MutSα/MutLα complexes displayed blinking by QD-MutLα (N=4/39); of these four nonblinking complexes, three were released from the mismatches upon the injection of ATP and began scanning the flanking DNA by 1D diffusion, and the fourth remained stationary at the lesion. Analysis of intersite transfer using the crisscrossed DNA curtains was confined only to those molecules of QD-MutSα, QD-MutLα, or MutSα/QD-MutLα complexes that exhibited QD blinking.

9. Spontaneous mismatch release by MutSα. QD-tagged proteins were bound to mismatch-bearing DNA in double-tethered curtains in buffer containing 20 mM Tris [pH 7.8], 50 mM NaCl, 1 mM ADP, 1 mM $MgCl_2$, 1 mM DTT, and 4 mg ml^{-1} BSA. The NaCl concentration was then raised to 150 mM to promote dissociation of nonspecifically bound proteins, and the remaining mismatch-bound molecules of MutSα were monitored continuously for a period of 10-15 minutes at an acquisition of 5 frames per second (200-msec integration). The QD-MutSα signals were then tracked, and escape from the lesions was defined as three contiguous frames outside the 3 standard deviations from the tracking noise; the probability of falsely identifying an escape event is on the order of $\sim 10^{-8}$. Of 76 proteins observed under these conditions, 53 remained bound to the lesions and did not escape the lesions (within experimental resolution as defined by 3 standard deviations from the tracking noise). The remaining 23 proteins showed clear 1D excursions away from the lesions, and of these 13 were analyzed by particle tracking, yielding a total of 95 lesion escape/return events; the remaining 10 proteins were not tracked because they collided with other nonspecifically bound proteins on the DNA. Analysis of the 95 excursions yielded a mean observed excursion distance and time of l_{obs}=3,134-bp and t_{obs}=30.7-seconds, and these values were in good agreement with theoretical expectations.

10. Protein dissociation with high salt chases. Proteins were bound to the DNA, and the number of proteins present was determined from the resulting images of the DNA. Video acquisition was then terminated, and 700-μl of reaction buffer containing either 300 mM NaCl (for MutSα) or 700 mM NaCl (for MutLα only or the MutSα/MutLα complex) was then flushed through the sample chamber at a flow rate of 0.2 ml min^{-1}. Images were collected from the same field, and used to determine the number of proteins that remained on the DNA after the high salt washes.

11. Intersite transfer assays. Crisscrossed curtains were made as described above for double-tethered curtains (8), with the exception that the DNA was sequentially injected from the two separate inlet channels (Fig. 6 & Fig. S10). The distance between the two DNA molecules was estimated by treating them as harmonic chains suspended above a reflective surface at a height equivalent to that of the nanofabricated barriers (Fig. S10).

Proteins were tracked in the absence of buffer flow. The tracking data was used to define the axes of the crisscrossed DNA molecules and the position of the intersection, and verified by staining with YOYO1. In cases where proteins showed obvious diffusion on the second DNA

molecule, the tracking data was fit to two lines using a least squares algorithm. When there was no apparent intersite transfer onto the second DNA, the positions of the two DNA molecules were defined separately. The first DNA molecule was defined by fitting the protein tracking data to a line; the second DNA molecule was defined by fitting the trace of another protein molecule that diffused on the second DNA molecule. The position of the intersection was calculated from the location of the lines. Tracking data were then centered so that the intersection was at $r_0(0, 0)$. Each data point was assigned to the intersection or one of the four arms of the intersection as follows:

1. Calculate the uncertainty of the intersection σ_{r0} based on the uncertainty of the two lines of DNA molecules.
2. Calculate the distance of each data point of the protein r_i to each arm d_1, d_2, d_3, and d_4.
3. Given the position and uncertainty of the intersection (r_0, σ_{r0}) and the data point (r_i, σ_{ri}) (σ_{ri} is the tracking precision of the protein), and a confidence level, judge if the data point is within the intersection. If so, assign the arbitrary designation '0' to this data point.
4. If the data point is outside the intersection area, assign it to be on one of the four DNA "arms" based the minimal value among d_1, d_2, d_3, and d_4. Assign '1', '2', '3' or '4' accordingly to this data point.

Intersite transfer was quantified based on the above assignment. If the tracking and assignment yielded the numeric sequence (1111110022033330111), where 1-3 correspond to different "arms" of the crisscrossed DNA relative to the DNA intersection, which is assigned as 0, we know that the protein underwent two intersite transfer events: from 1 to 0 to 2, and from 2 to 0 to 3, where "arm" 1 and 3 are on one DNA molecule and "arm" 2 and 4 are on another DNA molecule. Fig. 6g-j show color-coded examples of these assignments.

12. Origin crossings by a random walker. Fig. S6 shows the results from Monte Carlo simulations of a freely diffusing molecule with equally spaced absorbing boundaries (*e.g.*: nicks flanking either side of a mismatch). Boundary distances ranged from 2 to 190 steps away from the origin in the simulations, and 100,000 traces were generated for each boundary distance by selecting forward and backward steps with equal probability. The number of times the origin was encountered before the boundaries was recorded as well as the average number of steps necessary to encounter a boundary. As expected with a freely diffusing molecule, the average number of steps needed to travel a distance of N steps away was N^2 (Fig. S6a). Notably, the simulated traces also reveal that a molecule with equally spaced boundaries N steps away will on average cross the origin N-1 times (Fig. S6b).

This relationship can be further demonstrated by estimating the number of returns that occur in a given time (or steps), subject to particular boundary conditions. To estimate these values, we will use the conditional splitting probabilities, $\mathcal{E}_{n,m}(x)$, the conditional mean first passage time (CMFPT), $\tau_{n,m}(x)$, and a discrete crossing statistic, $C_m(L)$. $\mathcal{E}_{n,m}(x)$ is the probability that a walker starting at position x reaches a site located n away before reaching an opposing site m away. The splitting probabilities satisfy Laplace's equation, $\nabla_x^2 \mathcal{E}_{n,m}(x) = 0$, subject to the boundary conditions, $\mathcal{E}_{n,m}(n) = 1, \mathcal{E}_{n,m}(m) = 0$ (8), which when solved yields:

$$\mathcal{E}_{n,m}(x) = \frac{m - x}{m - n}$$

Similarly, $\tau_{n,m}(x)$ is the average time necessary for a walker starting at position x to reach a site located n away before reaching an opposing site m away. $\tau_{n,m}(x)$ can be obtained from Poisson's equation, $D_1\nabla_x^2[\mathcal{E}_{n,m}(x)\tau_{n,m}(x)] = -\mathcal{E}_{n,m}(x)$, subject to the boundary conditions, $\mathcal{E}_{n,m}(m)\tau_{n,m}(m) = 0, \mathcal{E}_{n,m}(n)\tau_{n,m}(n) = 0$, which yields:

$$\tau_{n,m}(x) = \frac{1}{6D_1}[(2m - n - x)(x - n)]$$

As an example, consider a walker beginning at the origin in between two absorbing boundaries located a distance N away. Once the walker takes a step away from the origin towards $\pm N$ it has a probability of N^{-1} to reach $\pm N$, i.e. $m = 0$, $n = \pm N$, and $x = \pm 1$. That is to say, the walker will, on average, fail to reach $\pm N$, $N - 1$ times before reaching it on the Nth try, and thereby return to the origin $N - 1$ times. Furthermore, during each failed excursion, a time $(N + 1)/3D_1$ elapses, and on the final successful excursion, $(N^2 + 2)/6D_1$ elapses, for an overall time for this process $\Delta t = N^2/2D_1$.

The final metric for redundancy in random walks is the number of crossing events, $C_m(N)$, at a particular location m, given a predetermined number of steps, L. For this calculation we describe the motion of a walker in discrete space and time. $C_m(L)$ is then related to the sum of the probability, $P(m, n)$, of the walker occupying the mth site at each step up to the nth step. $P(m, n)$ is given by the binomial distribution (9):

$$P(m, n) = \frac{2^{-n} n!}{\left(\frac{n+m}{2}\right)!\left(\frac{n-m}{2}\right)!}$$

If we are interested in the number of times a protein beginning at the origin will return to the origin in L steps, we simply sum $P(m, n)$ from 0 to L, allowing n to take on only even integers, due to the fact that only paths of even length can return to the origin, as given by:

$$C_0(L) = \sum_0^L P(0, n) \quad \text{for } n \in \text{evens}$$

$$C_0(L) = \frac{2}{\sqrt{\pi}} \frac{\Gamma\left(\frac{3+L}{2}\right)}{\Gamma\left(\frac{2+L}{2}\right)}$$

This relationship holds via simulation, confirming that at walker will cross its origin $C_0(L)$ times during a walk of L steps (Fig. S6c). Furthermore, to count the number of crossings at a

distal site a, $C_a(L)$, in L steps, we again count only the even paths to account for the fact that $P(a,n)$ includes both a and $-a$.

13. Spontaneous lesion escape and return. In modeling the transport process of MutSα on DNA, we first assume that the protein makes single base pair steps along the DNA. One of the consequences of probability density $P(m,n)$, is that the average displacement grows as the as the square root of the number of steps. That is to say, a protein takes on average 100 steps to reach an average distance of $10bp$ from its starting position. To connect this assumption with the experimental measurements, we impose an average stepping time, τ_{step}, such that after n steps a time $t = n\tau_{step}$ has elapsed.

Next, we designate two types of quantities: microscopic (experimentally observable within our resolution limits) and sub-microscopic (inferred from microscopic quantities and statistical analysis). Consider the following example to highlight these two suppositions. During an experiment, two successive measurements of position, $x_1(t_1)$ and $x_2(t_2)$, would yield a microscopic displacement $d = x_2 - x_1$ and a microscopic time step $\Delta t = t_2 - t_1$. From the above, we would then infer the protein made d^2 (or $\Delta t/\tau_{step}$) submicroscopic steps. Then using Einstein's relation for the mean squared displacement, we can relate τ_{step} to the microscopically measured diffusion coefficient as, $\tau_{step} = (2D)^{-1}$, when D and d^2 are given in identical units.

We define proteins that remain in the local environment of the lesion as microscopically bound (MB) to the mismatches. The local environment is defined as extending three standard deviations on either side of the lesion (840 bp; determined by examination of proteins stably bound to DNA, $\langle\sigma\rangle = 35nm \approx 140bp$). A protein was considered to have released the mismatch when three consecutive position measurements fell outside of the MB region. The probability of this observation in the event that the protein remained at the lesion is $\mathcal{O}(10^{-8})$.

We examined 95 events were MutS spontaneously dissociated from the mismatches and diffused a short distance along the DNA before rebinding. For each of these events we measured two microscopic values: the excursion length, l_{ex}, which is defined as the maximum distance that the protein travels away from the lesions during an excursion, and τ_{ex} which corresponds to total time that the protein spends diffusing on the DNA before rebinding to the lesions (Fig. S7). Given these definitions we can predict the probability of an excursion of length N from the splitting probabilities above as $P(l_{ex} = N) = \mathcal{E}_{N,0}(1)\mathcal{E}_{0,N+1}(N)$. This is the product of the probability to reach a site N, but not $N+1$, before returning to the origin. Inserting the solutions for the splitting probabilities from above, gives:

$$P(l_{ex} = N) = [N(N+1)]^{-1}$$

It is inaccurate to define an "average" theoretical length for this process because the distribution of excursions is divergent and can in principle extend to infinity. However, from the splitting probabilities we recognize that the probability of release from the mismatch and escaping the MB region without reencountering a lesion is only ~0.25%. Furthermore, of the proteins that escape the MB region, roughly 85% will return to the lesions before reaching a distance of 2.5-kb, and 90% of the total excursions are expected to be 10-bp or less (Fig. S7) Thus the majority of escape/return events should be occurring within the submicroscopic regime and thus would not be detectable in our microscopic observations.

We can also calculate the mean time associated with an excursion of length N by adding the CMFPT to reach N from the origin to the CMFPT to reach the origin from N; $\tau_{ex}(l_{ex} = N) = \tau_{N,0}(1) + \tau_{0,N+1}(N)$, which gives:

$$\tau_{ex}(l_{ex} = N) = \frac{2N(N+1) - 1}{6D_1}$$

This value includes time spent within the MB region, therefore, we use a modified excursion time, $\tilde{\tau}_{ex}(N > 420)$, which is the time of the excursion spent outside of the MB region given as:

$$\tilde{\tau}_{ex} = \tau_{ex}(N) - \tau_{ex}(420bp)$$

From this, we can show that of the proteins that escape the MB region, ~50% will spend 1 second or less away from the MB region and the same 85% from above will spend fewer than 13 seconds outside the MB region before rebinding the lesions. As expected, 90% of the total population will spend fewer than $250\,\mu s$ away from the lesion and would fall outside the microscopically observable regime.

14. Apparent versus direct 3D binding. A protein was categorized as having undergone a 1D search only if there were at least two frames at the beginning of the diffusion trajectory that were at least three standard deviations away from the location of the mismatch. If the proteins initially appeared within this resolution limit, then they were categorized as having undergone an apparent direct 3D binding event. Events categorized as apparent 3D binding could be attributed to 1D sliding on a submicroscopic scale, and this is likely true for many of these events. Therefore we sought to estimate what fraction of events ascribed to apparent 3D binding could be occurring through submicroscopic sliding and what fraction were likely to occur through direct 3D binding in the absence of submicroscopic 1D diffusion. To estimate the distribution of possible lengths, $P(\,l_{en})$, which result in target binding, we calculate the probability that a protein initially bound at site x_0 on the DNA finds a target (located at the origin) before dissociation. This probability, $P(x_0)$, is the product of two densities: the probability that a protein, which binds the DNA at time, $t = 0$, is still bound at a later time t, $\exp(-k_d t)$, where k_d defines the lifetime of the nonspecifically bound protein, and the probability that a protein starting at site, x_0, at time zero, encounters the mismatch for the first time at time t, $j(0, t|x_0, 0)$ (10).

Here $j(0, t|x_0, 0)$, is the concentration flux to the origin, provided the origin (mismatch) is an absorbing boundary, $C(0, t) = 0$, and the end of the DNA at $x = L$ acts as a reflecting boundary, $\frac{d}{dx}C(x, t)\Big|_{x=L} = 0$ (11). We then formulate $P(x_0)$ as the integral of the conditional probability of the protein finding the mismatch before dissociation over all time, yielding:

$$P(x_0) = \int_0^\infty dt \exp(-k_d t)\; j(0, t|x_0, 0)$$

$$P(x_0) = \cosh\left((x_0 - L)\sqrt{\frac{k_d}{D_1}}\right)\operatorname{sech}\left(L\sqrt{\frac{k_d}{D_1}}\right)$$

At low concentrations, there is likely only one protein searching the DNA at a time. This protein has an equal probability of starting its search at any site on the DNA. Therefore, we calculate the probability of observing a particular encounter length, l_{en} as the product of the probability of landing at a particular site and $P(x_0)$. Where the normalization expresses the fact that eventually (potentially after dissociation and rebinding) the target will be found.

$$P(\,l_{en}) = \frac{P(x_0)}{L\int_0^L P(x_0)\,dx_0} = \sqrt{\frac{k_d}{D_1}}\cosh\left((L-x_0)\sqrt{\frac{k_d}{D_1}}\right)\operatorname{csch}\left(L\sqrt{\frac{k_d}{D_1}}\right)$$

We can then estimate the probabilities of apparent 3D binding events, $P(3d)$, and observable 1D binding events, $P(1d)$, as:

$$P(1d) = \sinh\left((L-a)\sqrt{\frac{k_d}{D_1}}\right)\operatorname{csch}\left(L\sqrt{\frac{k_d}{D_1}}\right)$$

$$P(3d) = 1 - \sinh\left((L-a)\sqrt{\frac{k_d}{D_1}}\right)\operatorname{csch}\left(L\sqrt{\frac{k_d}{D_1}}\right)$$

Based on current spatial resolution limits, and our categorization of target binding based on three standard deviations from the mismatch, these calculations predict that ~80% of the experimentally observed events would have been categorized as occurring through 1D sliding, and the remaining ~20% of observed events would have been categorized as 3D collisions; these results would have been the same for both MutSα and MutLα because both proteins have relatively long lifetimes (≥20 sec) on nonspecific DNA. The experimental data do not reflect these predicted distributions, rather 57.5% and 45% of the experimentally observed target binding events were categorized as direct 3D target binding for MutSα and MutLα, respectively. As a simple approximation, the discrepancy between the model and the experimental results suggest that ~38% of MutSα targeting events and ~25% of MutLα targeting events can be explained by direct 3D binding in the absence of any submicroscopic 1D sliding.

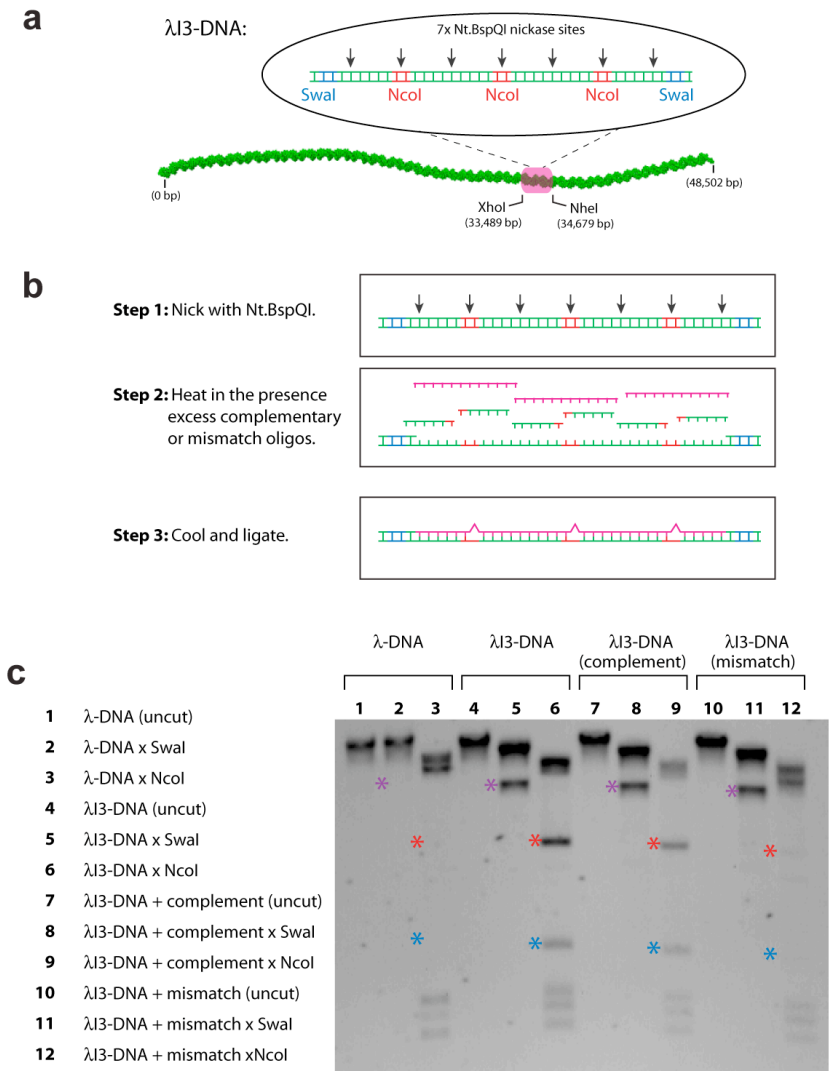

Fig. S1. Construction and characterization of mismatch substrate. (**a**) Overview of λ I3-DNA construction. Highlights sites used for inserting the new DNA fragments, as well as the number and arrangement of the Nt. BspQI nickase sites, and restriction sites for NcoI and SwaI. (**b**) Schematic of oligonucleotide insertion strategy. The **λ**I3-DNA is green, and the nicking sites are indicated with arrowheads. After treatment with Nt. BspQI , the λ-DNA is mixed with an excess of the appropriate oligonucleotide (magenta), and then briefly heated and cooled to replace the nicked fragments with the new oligonucleotide. (**c**) Restriction analysis of the **λ**I3-DNA substrates. The **λ**I3-DNA substrate has three unique restriction fragments not present in the wt phage: digestion with SwaI yields a 13,828 bp fragment (purple asterisk) and digestion with NcoI liberates two fragments 9,959 bp (red asterisk) and 5,671 bp (blue asterisk) in length. Insertion of the mismatch eliminates the two fragments produced by digestion with NcoI (because the mismatch disrupts the NcoI site), but does not affect the SwaI fragment.

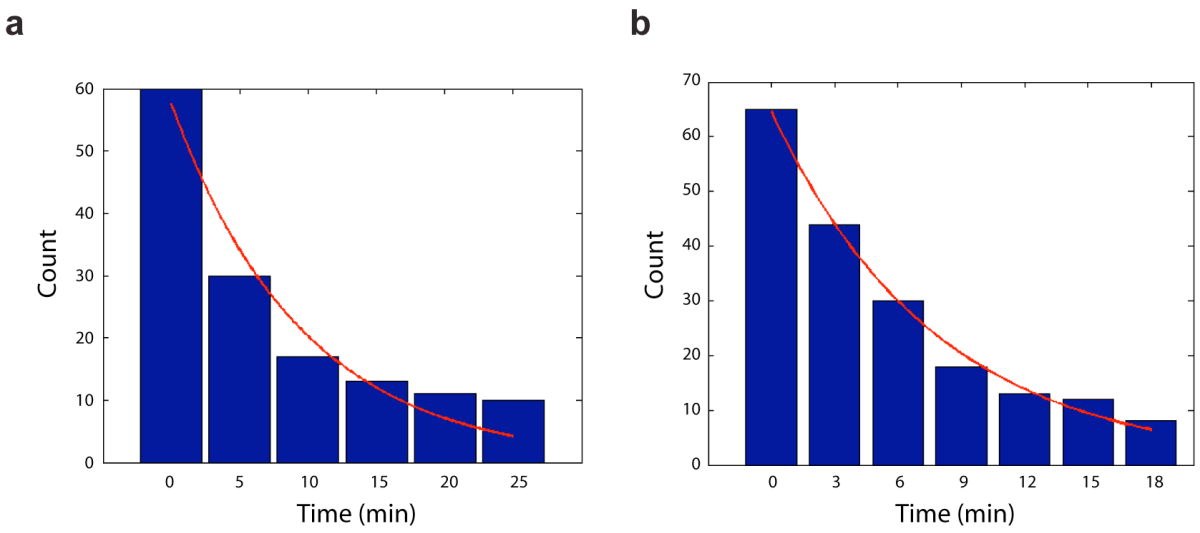

Fig. S2. Half-life of MutSα bound at mismatches and MutLα bound to the MutSα-mismatch complex. (**a**) QD-tagged MutSα was bound to the 3x mismatches on a single-tethered DNA curtain as shown in Fig. 1 of the main text. The lesion-bound proteins were then chased with buffer that lacked additional free protein and contained 150 mM NaCl and 1 mM ADP (along with 20 mM Tris [pH 7.8], 1 mM $MgCl_2$, 1 mM DTT, and 4 mg ml^{-1} BSA). The number of proteins that remained bound to the mismatch was measured at defined time intervals, and the resulting data were fit with single exponential curves to determine the half-lives of the lesion-bound proteins, yielding a value of 9.6±1.5 minutes for lesion-bound MutSα. (**b**) QD-tagged MutLα was bound to the mismatch-MutSα (untagged) complex as shown in Fig. 4d of the main text. The lesion-bound MMR proteins were then chased with buffer that lacked additional free protein and contained 150 mM NaCl and 1 mM ADP (along with 20 mM Tris [pH 7.8], 1 mM $MgCl_2$, 1 mM DTT, and 4 mg ml^{-1} BSA). The number of QD-tagged MutLα proteins that remained bound to the DNA was measured at defined time intervals, and the resulting data were fit with single exponential curves to determine the half-lives of the lesion-bound proteins, yielding a value of 7.8±0.4 minutes.

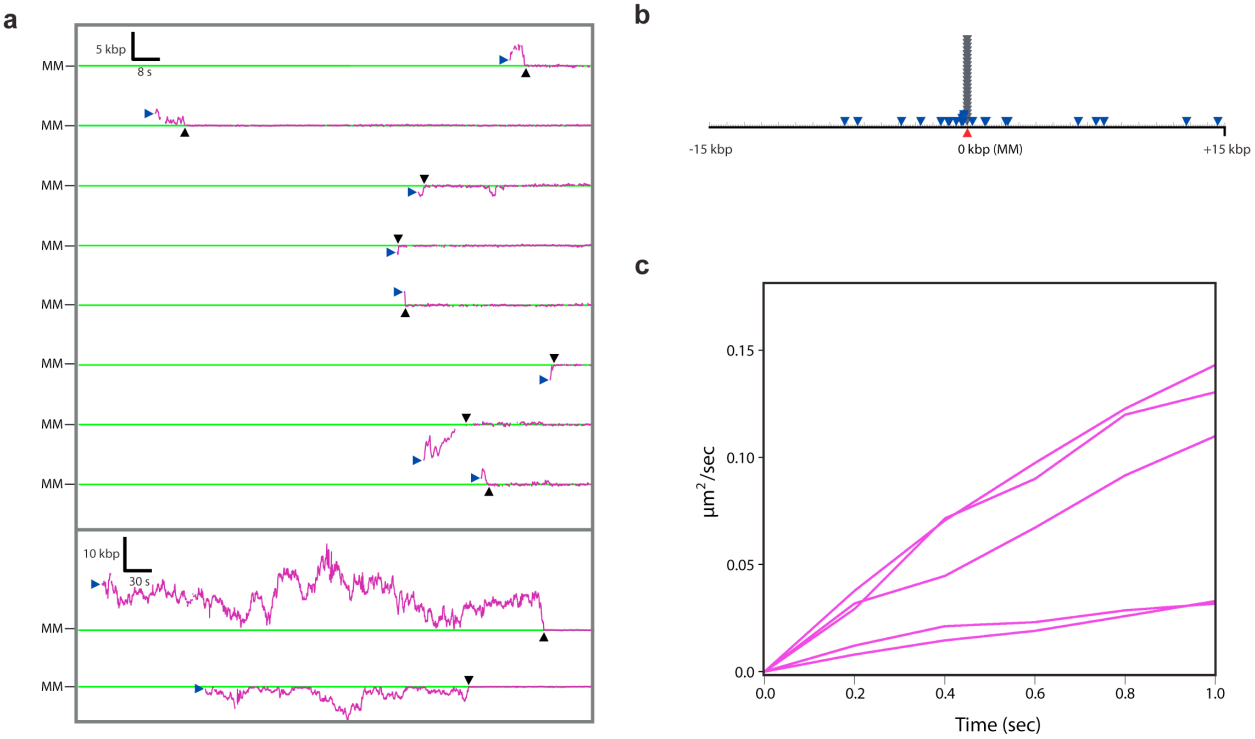

Fig. S3. Mismatch targeting by MutSα. (**a**) Shows 10 representative examples of tracking data (magenta) for molecules of MutSα that engaged the mismatches through a 1D search. The initial binding positions of the proteins are indicated with blue arrowheads, the location of the mismatches is indicated as MM and a green line, and lesion engagement is indicated with black arrowheads. (**b**) Shows a map of the initial binding sites for all observed molecules of MutSα that bound to the lesions. Gray arrow heads correspond to proteins that bound directly to the mismatches through an apparent 3D mechanism (within our optical resolution limits) and blue arrowheads correspond to proteins that bound to nonspecific DNA sites and slid in 1D along the DNA to engage the lesions. (**c**) Shown are five representative examples of MSD plots generated from the tracking data of MutSα as it searches for lesions.

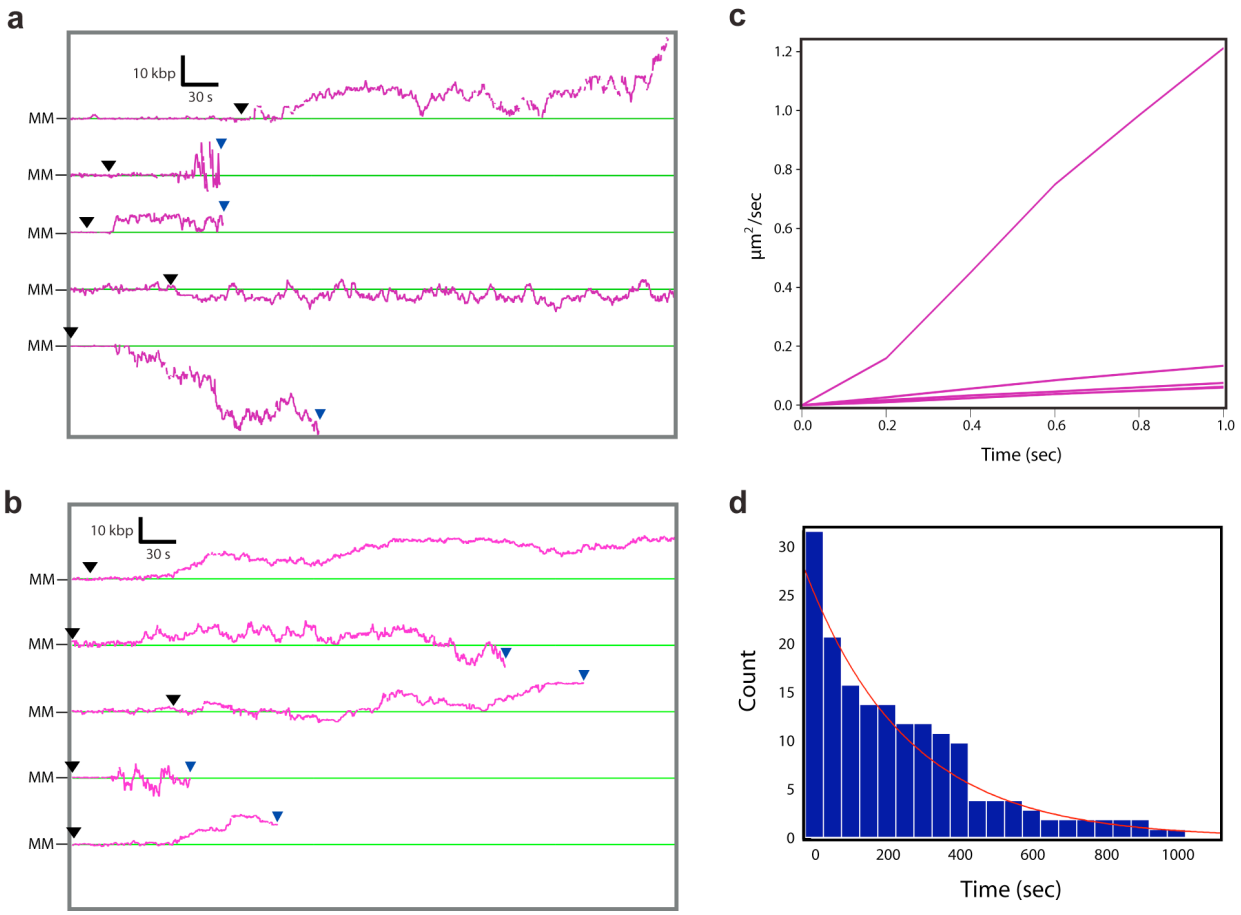

Fig. S4. 1D diffusion of mismatch-bound MutSα following ATP or ATPγS chase. (**a**) Shows 5 representative examples of tracking data illustrating the behavior of mismatch bound MutSα after the injection of 1 mM ATP, and (**b**) shows 5 representative examples of tracking data illustrating the behavior of mismatch bound MutSα after the injection of 1 mM ATPγS. Gaps in the tracking data correspond to portions of the trajectories that could not be accurately tracked due to QD blinking or changes in background intensity, and the end-points in the tracking data correspond to dissociation of the proteins from the DNA. In both (**a**) and (**b**), black arrowheads indicate when ATP or ATPγS enters the microfluidic sample chamber, blue arrowheads indicate when the protein dissociates from the DNA, and for traces lacking blue arrowheads the proteins remained bound to the DNA and continued diffusing beyond the data collection window. (**c**) Shown are five representative examples of MSD plots generated from the tracking data of MutSα after being released from lesions upon chasing with ATP. (**d**) Distribution of lifetimes for MutSα/MutLα after being released from the mismatches upon ATP injection (N=32). These values yield a lower bound on the lifetime of the diffusing MutSα/MutLα complex after ATP-triggered release from the lesions of $t_{1/2} \geq 198 \pm 23.4$ seconds. Note that this value is a lower bound because ~40% of the observed molecules did not dissociate from the DNA during the observation windows, rather they remained bound to the DNA and kept diffusing.

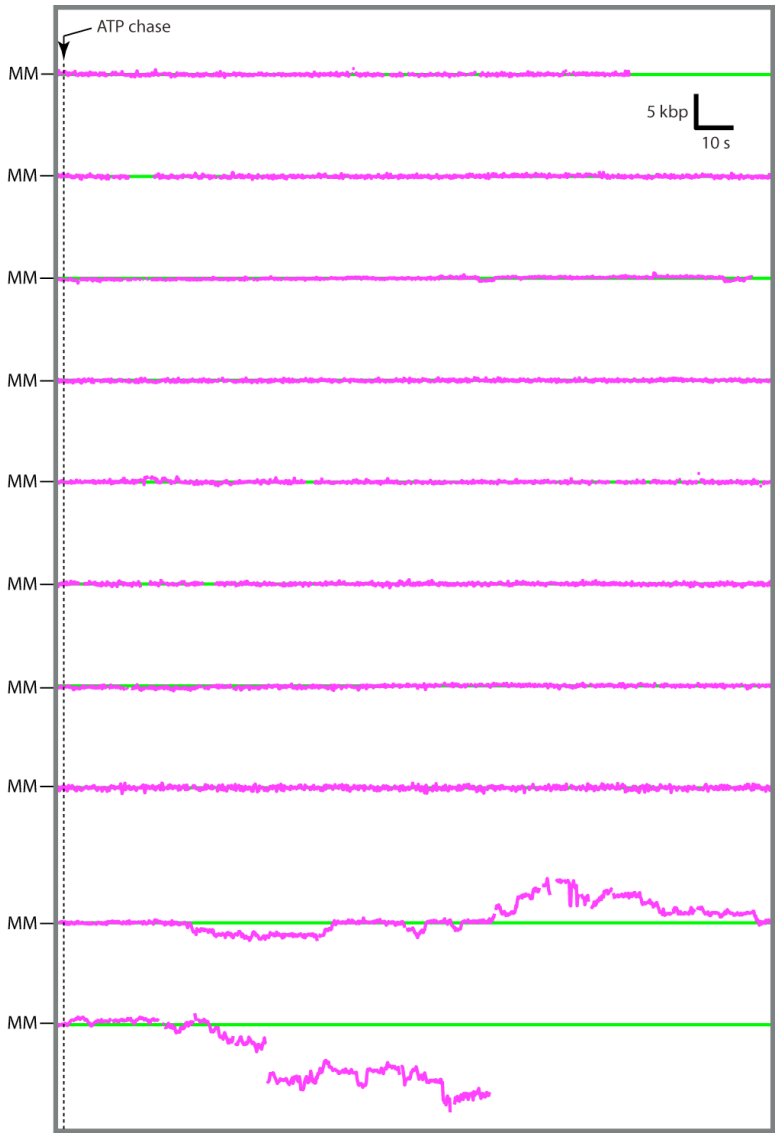

Fig. S5. ATP chase of mismatch-bound MutSα in low salt buffer. Data were collected in 1 mM ADP and 50 mM NaCl (along with 20 mM Tris [pH 7.8], 1 mM $MgCl_2$, 1 mM DTT, and 4 mg ml^{-1} BSA), and then chasing the bound proteins with the same buffer with the exception that 1 mM ADP was replaced with 1 mM ATP, as indicated by the dashed line. Upon chasing with ATP at 50 mM NaCl, 18% of the proteins began diffusing along the DNA, 4% directly dissociated, and 78% remained stationary at the lesions (N=78). The tracking data above highlight ten representative examples of different MutSα molecules, including 8 that remained stationary and 2 that exhibited 1D excursions away from the mismatches. Notably, one of the two molecules that escaped from the mismatch also rebound to the mismatch following brief 1D excursions (second trace from the bottom). This is the only example of lesion rebinding by MutSα that we have seen in the presence of ATP, and could suggest that the complexes remain in the ADP bound state under these low salt conditions; however, we are hesitant to draw conclusions given the rarity of this observation.

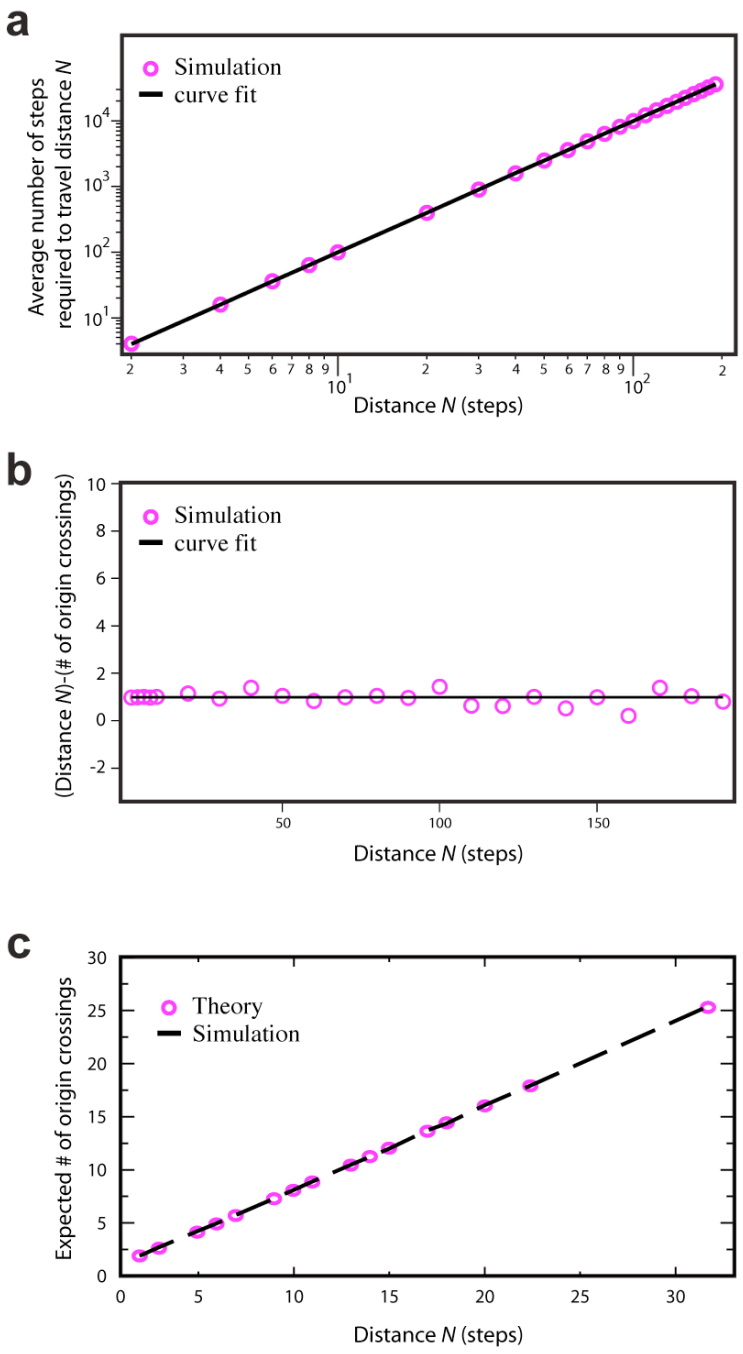

Fig. S6. Redundant nature of 1D-diffusion. (**a-b**) Results of simulations of a 1D random walk, which were used to reveal the average number of steps (*N*) necessary to move a given distance along a 1D lattice (**a**), and difference between the number of steps (*N*) taken and the average number of origin (or mismatch) crossings (**b**). (**c**) Theoretical number of origin crossings versus the number of steps taken.

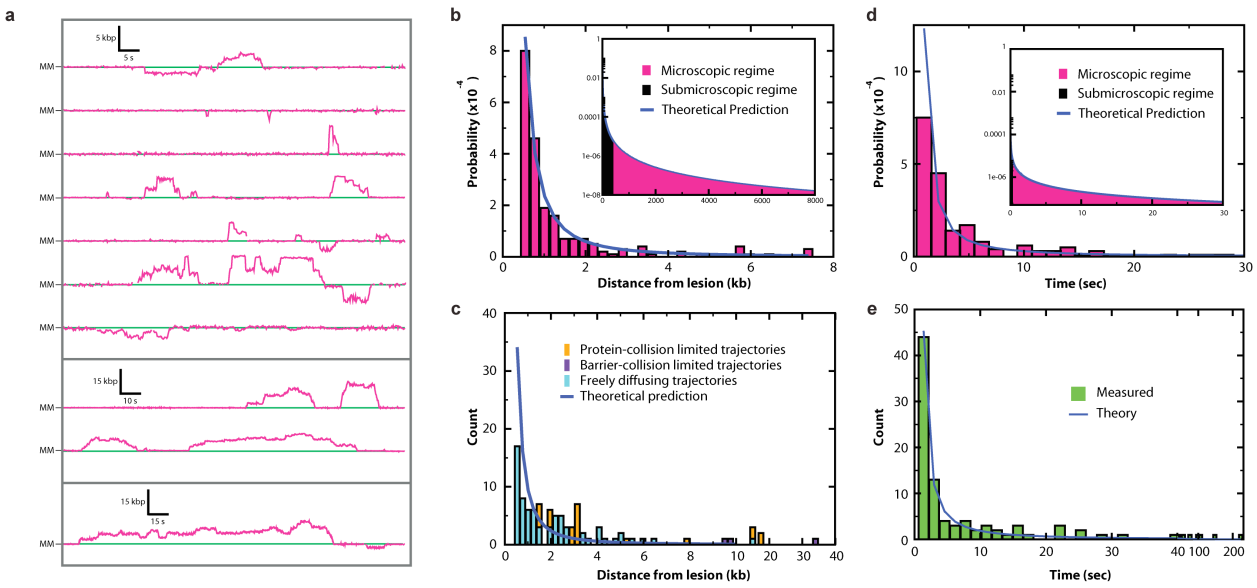

Fig. S7. Spontaneous mismatch escape and return by MutSα. Data were collected by monitoring lesion-bound MutSα in buffer containing 1 mM ADP and 150 mM NaCl (along with 20 mM Tris [pH 7.8], 1 mM $MgCl_2$, 1 mM DTT, and 4 mg ml^{-1} BSA). The proteins were continuously observed for at ten minutes at 10 Hz. The tracking data highlight examples of different MutSα molecules that spontaneously escaped from the mismatched bases. The MutSα trajectories are shown in magenta, and the location of the mismatches (MM) is indicated by the green line. (**b**) Inset, theoretical calculation for the sliding distance after spontaneous mismatch release based on 1D random walk, revealing a mean excursion distance of l_{theor} = 22.5-bp and a mean return time of t_{theor}= 1.5 milliseconds. Results are segregated into microscopically observable (magenta) and submicroscopic regimes (black), and the blue line represents a fit to the calculations. Larger graph shows the theoretical expectation based on simulations of 10,000 random walkers and the resulting data are displayed at the resolution of our experimental data (see below), yielding mean observable excursion distances and times of $l_{theor,obs}$=2,014-bp and $t_{theor,obs}$=11.5-seconds; the submicroscopic regime is omitted from this plot because it would not be experimentally observable. (**c**) Experimental observations. The experimentally observed diffusion trajectories are segregated into proteins that freely diffused on the DNA (cyan), and those whose diffusion distance was limited by collisions with either other proteins on the DNA (yellow) or the chromium barriers (purple). The experimental data yielded observed excursion distances and times of l_{obs}=3,134-bp and t_{obs}=30.7-seconds, which are in good agreement with the theoretical expectations. (**d**) Inset, theoretical calculation for the return time after spontaneous mismatch release based on 1D random walk. The results are segregated into microscopically observable regime (magenta), the submicroscopic regime (black), and the blue line represents a fit to the calculations and is the same in all three graphs. Larger graph shows the theoretical expectation and the resulting data are displayed at the resolution of our experimental data; the submicroscopic regime is omitted because it would not be experimentally observable. (**e**) Experimental observed return times.

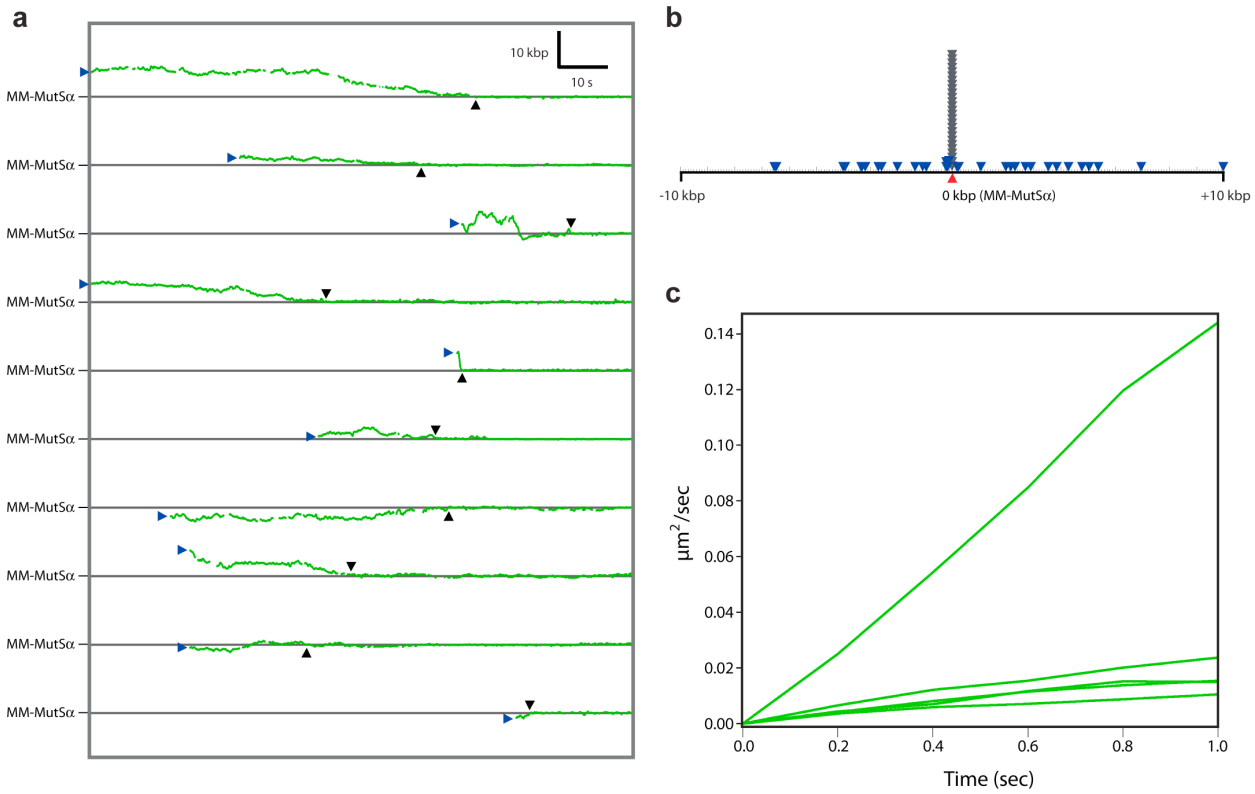

Fig. S8. Mismatch-MutSα targeting by MutLα. (**a**) Shows 10 representative examples of tracking data for molecules of MutLα that engaged mismatch-bound MutSα through a 1D search. The initial binding positions of the proteins are indicated with blue arrowheads, the locations of the MutSα-mismatches are indicated, and lesion engagement is indicated with black arrowheads. (**b**) Shows a map of the initial binding sites for all observed molecules of MutLα that bound to mismatch-bound MutSα. Gray arrow heads correspond to MutLα proteins that bound directly by apparent 3D collisions to the mismatch-bound MutSα (within optical resolution limits) and blue arrowheads correspond to MutLα proteins that bound to nonspecific DNA sites and slid in 1D along the DNA to engage the mismatch-bound MutSα.
(**c**) Five representative examples of MSD plots generated from the tracking data of MutLα as it searches for lesion-bound MutSα.

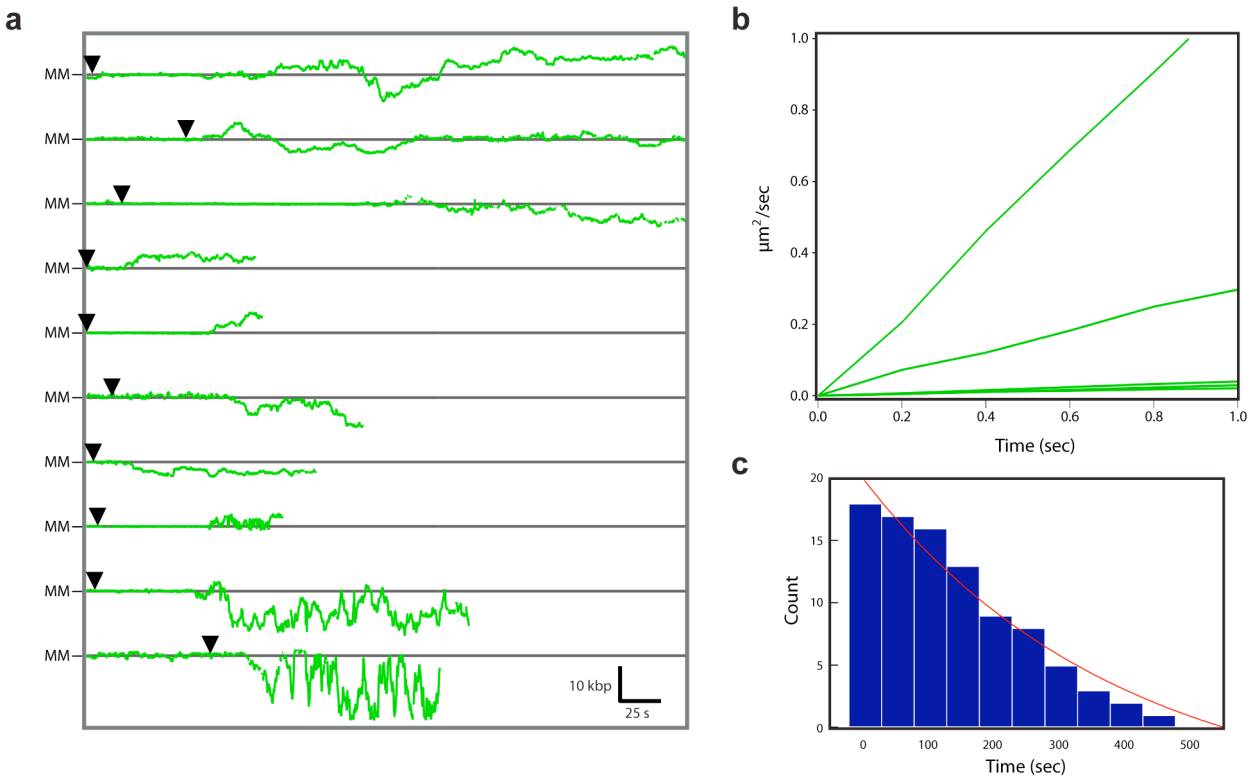

Fig. S9. ATP-triggered release of the mismatch-bound MutSα/MutLα complex. (**a**) Ten representative examples of tracking data illustrating the behavior of mismatch bound MutSα after injection of ATP, and including complexes that exhibited blinking of QD-MutLα (top eight traces) as well as nonblinking QD-MutLα (bottom two traces; QD-MutS exhibit blinking in all observed cases). The black arrowheads indicate when ATP entered the flowcells. Gaps in the tracking data correspond to portions of the trajectories that could not be accurately tracked due to QD blinking or changes in background intensity. The ends of the traces correspond to the end of the data collection, and do not correspond to protein dissociation from the DNA. (**b**) Five representative MSD plots for the MutS/MutL complex after ATP-triggered lesion release. (**c**) The distribution of lifetimes measured for MutSα/MutLα after being released from the mismatches upon ATP injection (N=18). These values yield a lower bound on the lifetime of the diffusing MutSα/MutLα complex after ATP-triggered release from the lesions of $t_{1/2} \geq 267.6 \pm 62.1$ seconds. Note that this value is a lower bound because ~70% of the observed complexes did not dissociate from the DNA during the observation windows, rather they remained bound to the DNA and kept diffusing, therefore they are not included in the histogram.

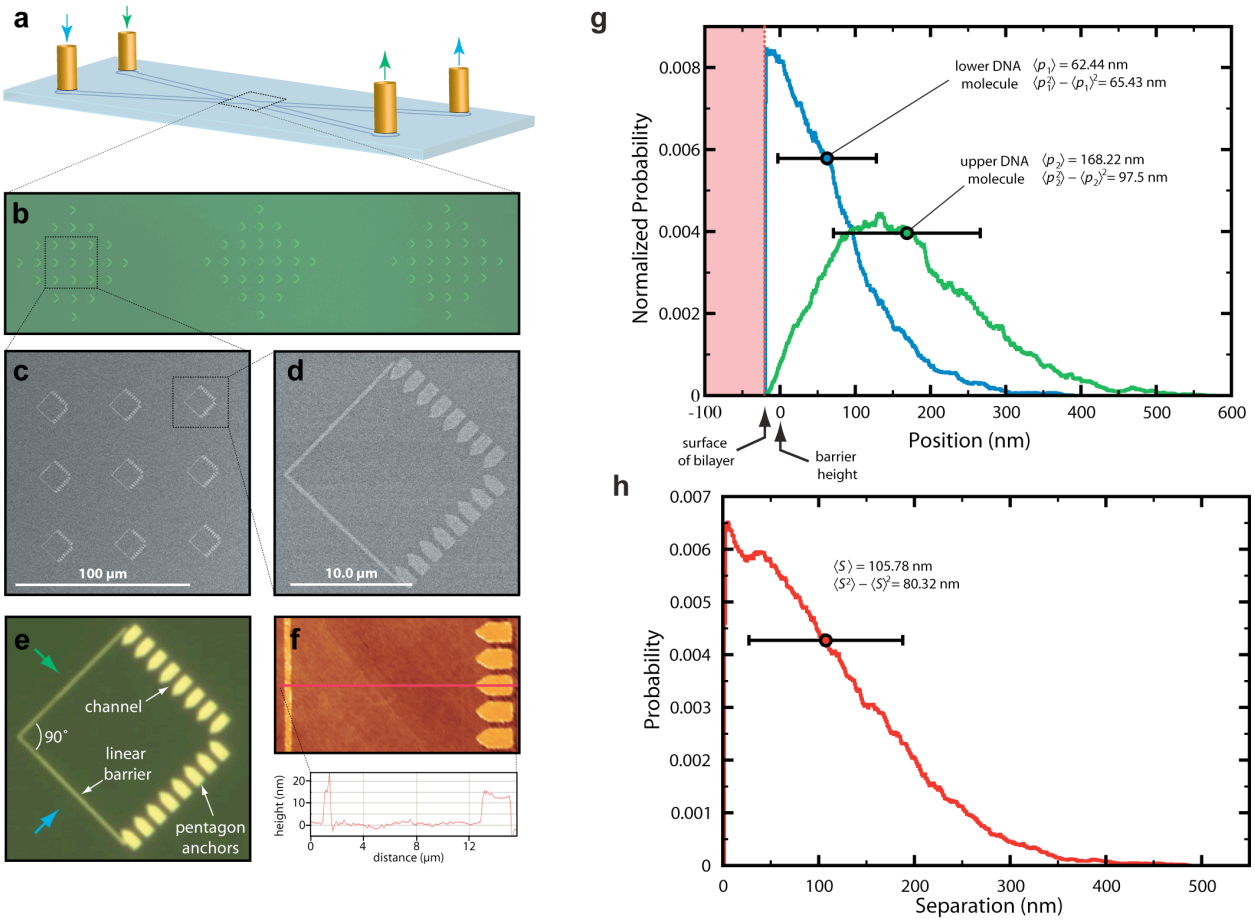

Fig. S10. Barrier patterns for crisscrossed DNA curtains. (**a**) Schematic of the two-channel flowcell. (**b**) Low magnification (10x) optical image of the chromium (Cr) patterned surface. (**c** and **d**) High-resolution SEM (scanning electron microscope) images of the Cr patterns. (**e**) High magnification (100x) optical image of a single pattern. Important elements of the pattern design are highlighted. (**f**) AFM (atomic force microscope) image illustrating barrier height. (**g**) This graph shows the calculated distance between two crisscrossed DNA molecules suspended 20-nm above a solid surface (represented in red) as determined from a 20-minute simulation with 200-nsec time steps. The position for the lower DNA is shown in blue, and the upper DNA is shown in green. The average positions $\langle p_i \rangle$ of the DNA molecules relative to the bilayer surface, and dispersion $\langle p_i^2 \rangle - \langle p_i \rangle^2$ about these positions, are indicated in (**a**). The average positions of the DNA relative to one another $\langle S \rangle$, and the corresponding dispersion $\langle S^2 \rangle - \langle S \rangle^2$ about these positions, are shown in (**h**).

Supplemental References

1. Gorman J, et al. (2007) Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6. *Mol Cell* 28(3):359 - 370.
2. Gorman J, Plys A, Visnapuu M, Alani E, & Greene E (2010) Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat Struct Mol Biol* 17(8):932 - 938.
3. Visnapuu M-L & Greene E (2009) Single-molecule imaging of DNA curtains reveals intrinsic energy landscapes for nucleosome deposition. *Nat Struct Mol Biol* 16:1056-1062.
4. Efron B & Tibshirani R (1993) An Introduction to the Bootstrap (Champman and Hall, Inc., New York).
5. Dahan M, et al. (2003) Diffusion dynamics of glycine receptors revealed by single-quantum dot tracking. *Science* 302(5644):442 - 445.
6. Yao J, Larson D, Vishwasrao H, Zipfel W, & Webb W (2005) Blinking and nonradiant dark fraction of water-soluble quantum dots in aqueous solution. *Proc Natl Acad Sci U S A* 102(40):14284 - 14289.
7. Zhang Q, Li Y, & Tsien R (2009) The dynamic control of kiss-and-run and vesicular reuse probed with single nanoparticles. *Science* 323(5920):1448 - 1453.
8. Gorman J, Fazio T, Wang F, Wind S, & Greene E (2010) Nanofabricated racks of aligned and anchored DNA substrates for single-molecule imaging. *Langmuir* 26:1372 - 1379.
9. Feller W (1971) An introduction to probability theory and its applications (John Wiley & Sons, Inc.).
10. Van Kampen N (2007) Stochastical Processes in Physics and Chemistry (Elsevier Press).
11. Redner S (2001) A guide to first passage processes (Cambridge University Press).