SUPPORTING INFORMATION

for

Dynamic human MutSα-MutLα complexes compact mismatched DNA

Kira C. Bradford[1,6†], Hunter Wilkins[1†], Pengyu Hao[5], Zimeng Li[3], Bangchen Wang[1], Dan Burke[2], Dong Wu[1], Austin E. Smith[1], Logan Spaller[1], Chunwei Du[4], Jacob W. Gauer[1], Edward Chan[5], Peggy Hsieh[4], Keith R. Weninger[5*], Dorothy A. Erie[1,3,*]

[1]Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599

[2]Department of Physics, University of North Carolina, Chapel Hill, NC 27599

[3]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599

[4]Genetics and Biochemistry Branch, NIDDK, National Institutes of Health, Bethesda, MD 20892

[5]Department of Physics, North Carolina State University, Raleigh, NC 27695

[6] Renaissance Computing Institute, University of North Carolina, Chapel Hill, NC 27599

*Correspondence to: derie@unc.edu or keith.weninger@ncsu.edu

†These authors have contributed equally to the work.

Includes:
Supporting table S1
Supporting figures S1-S6
Supporting Methods

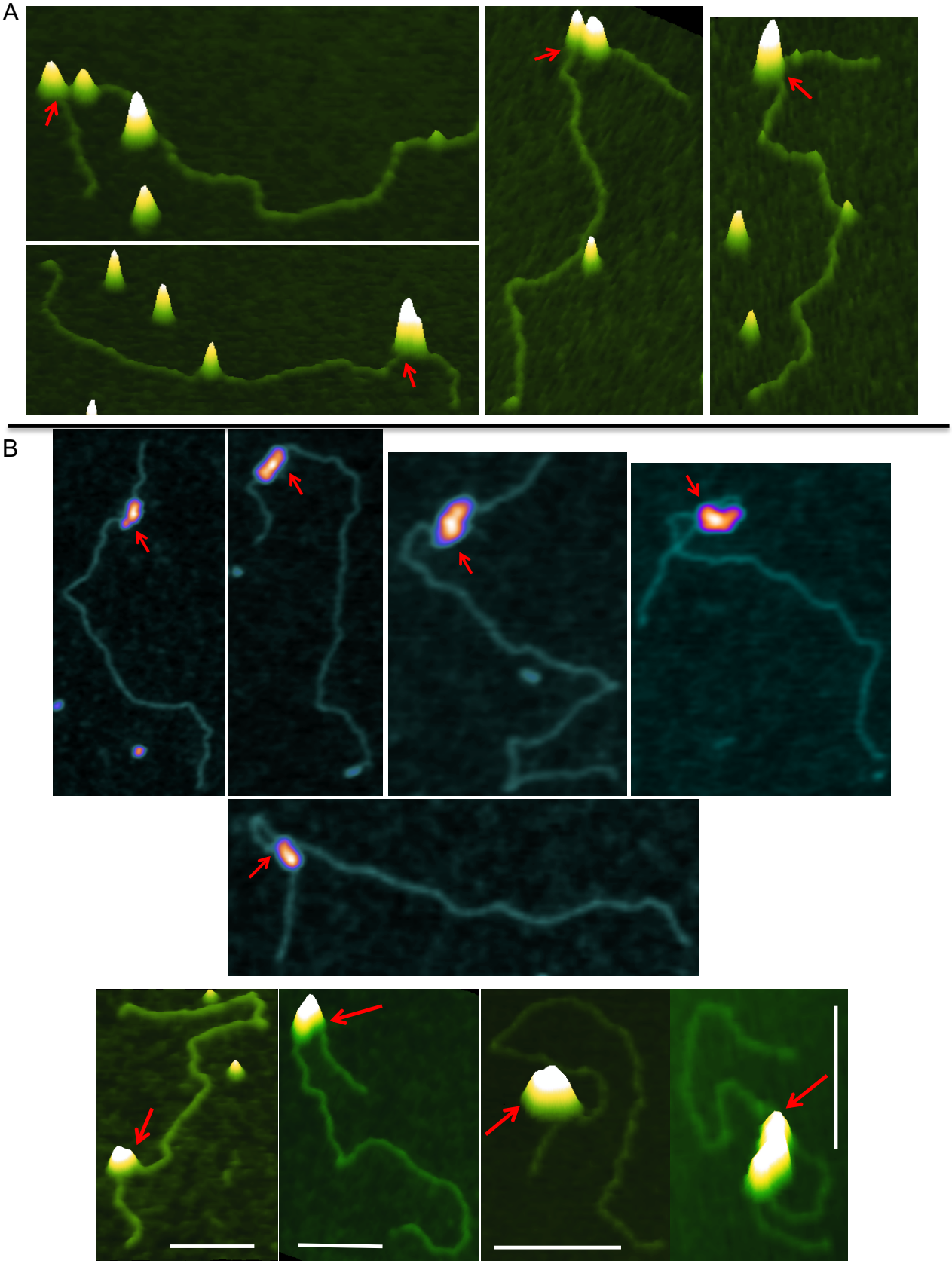

Figure S1. Additional AFM images of MutSα- and MutSα-MutLα-GT-DNA complexes. (A) Examples showing 3-D topographic views of MutSα on GT-DNA with ATP. **(B)** Examples of MutSα-MutLα complexes on GT-DNA that involve the mismatch. Arrows point to the mismatch.

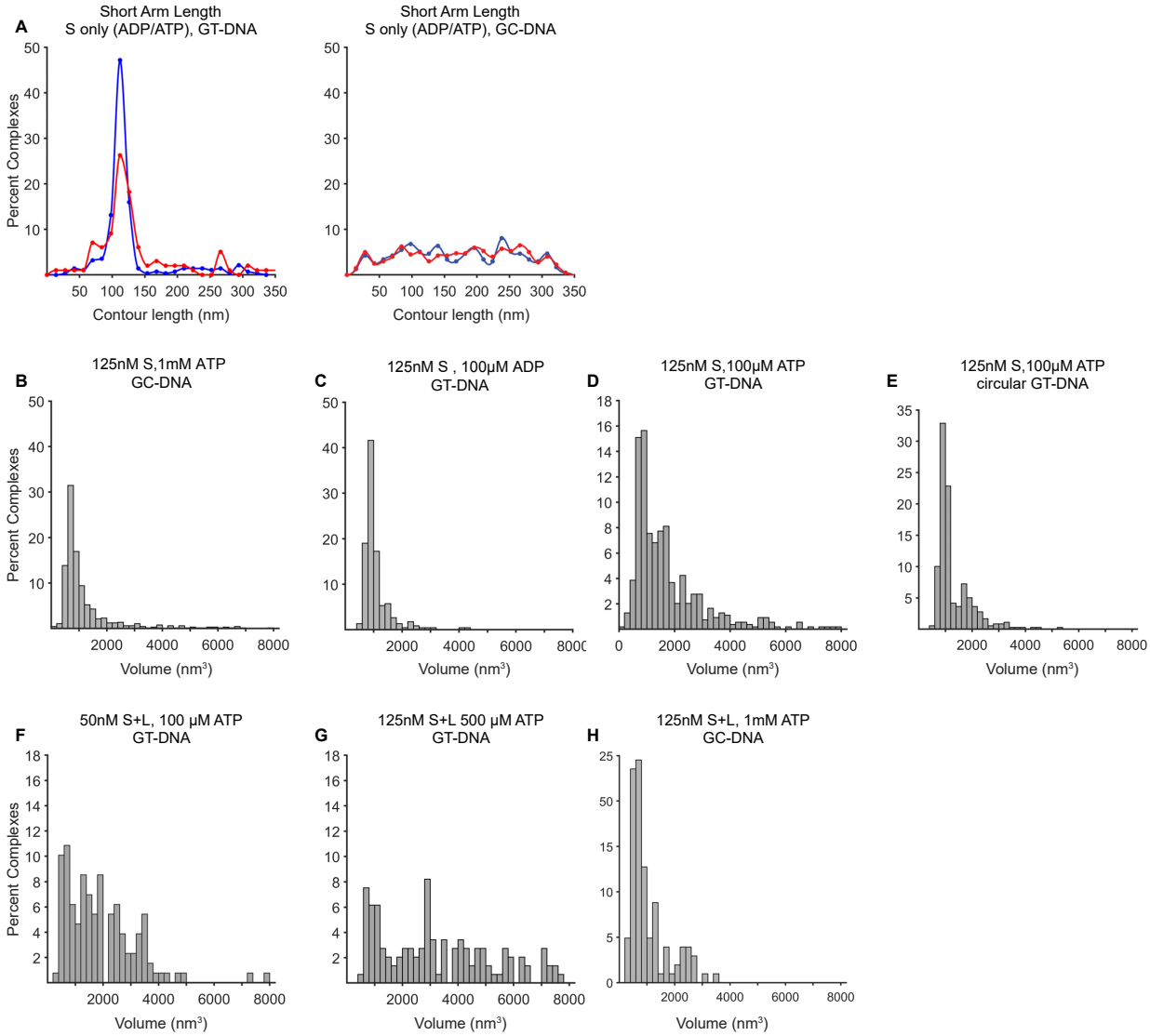

Figure S2. Position, length, and volume analysis of MutSα- and MutSα-MutLα-DNA complexes. **(A)** Distributions of the position of MutSα on the DNA (short arm length) in the presence of 100 µM ADP (blue; n=282 for GT, n=310 for GC) or 1 mM ATP (red; n=256 for GT, n=401 for GC) on GT-DNA (Left) and GC-DNA (Right). **(B-E)** Distribution of volumes of MutSα-DNA complexes for 125 nM MutSα incubated with **(B)** 1 mM ATP and GC-DNA for 2 min (n=649), **(C)** 100 µM ADP and GT-DNA for 1 min (n=226), **(D)** 100 µM ATP and GT-DNA for 1 min (n=543), **(E)** 100 µM ATP and nicked plasmid (circular) GT-DNA for 1 min (n=359). **(F-G)** Distributions of volumes of protein-GT-DNA complexes for **(F)** 50 nM each of MutSα + MutLα incubated with100 µM ATP for 1 min (n=129) and **(G)** 125 nM each of MutSα + MutLα incubated with 500 µM ATP for 1 min (n=146). **(H)** Distributions of volumes of protein-GC-DNA complexes for 125 nM each of MutSα + MutLα incubated with 1 mM ATP for 5 min (n=102). These results demonstrate that the larger MutSα-MutLα-DNA complexes require a GT-mismatch and that our crosslinking protocol does not result in formation of nonspecific SL complexes in the absence of a mismatch.

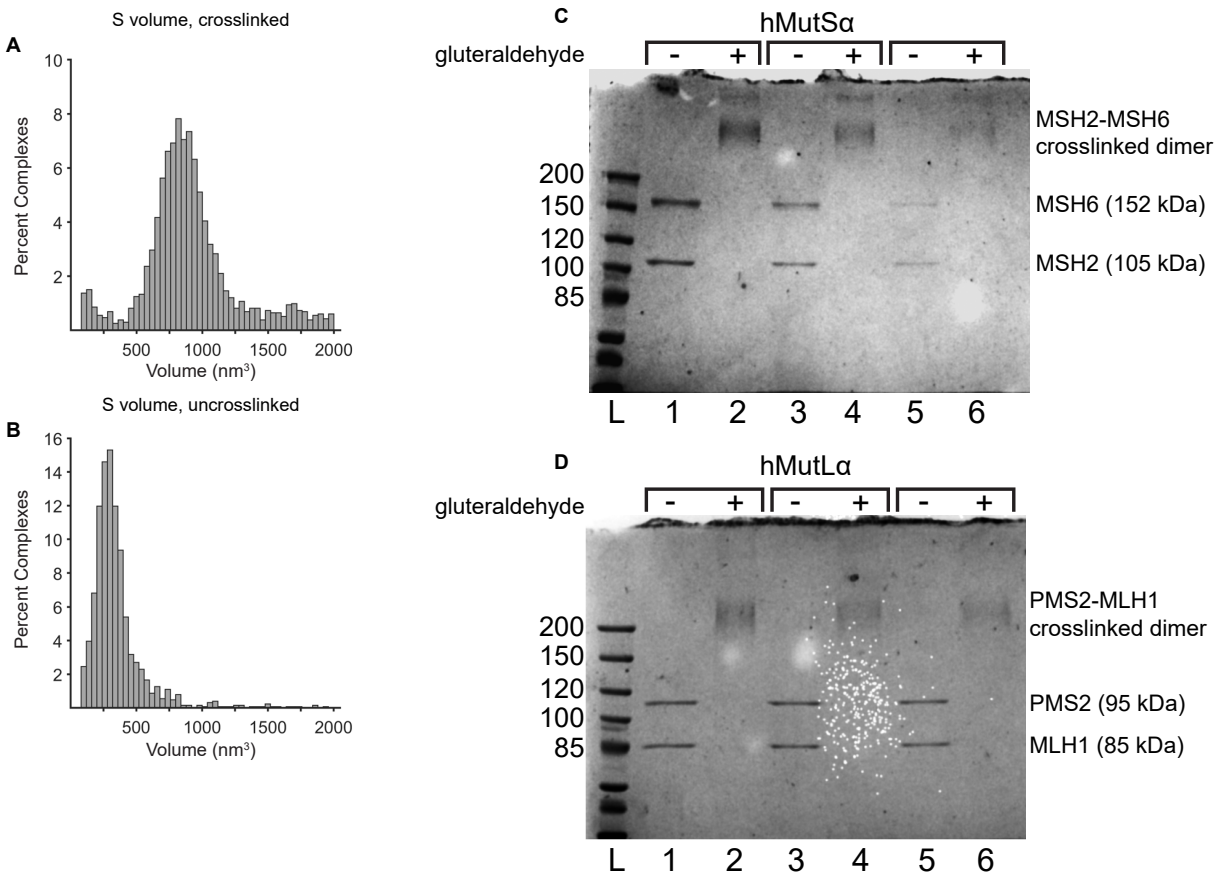

Figure S3. Characterization of crosslinking. (A-B) Volume distributions of free MutSα in the presence of ATP when **(A)** crosslinked (n=2327) or **(B)** uncrosslinked (n=1261). The increased AFM volume of crosslinked MutSα results primarily from an increase in the height of crosslinked MutSα (average height ~ 2.6 nm) relative to uncrosslinked MutSα (average height ~ 0.9 nm). Crosslinking stabilizes the heterodimer (C), which, in turn, reduces the interaction the mica surface resulting in the protein "sitting" taller on the surface. The dependence of height on the measured volume in AFM images is complex because the finite size of the tip (~5-7 nm radius of curvature) leads to a dilation of the features in the image (1). **(C-D)** Six percent SDS-polyacrylamide gels showing uncrosslinked controls and crosslinked products for 125 nM MutSα **(C)** or MutLα **(D)** in the presence of ATP (1mM) (lanes 1 & 2), ATP + GC-DNA (lanes 3 & 4), or ATP + GT-DNA (lanes 5 & 6). The absence or presence of glutaraldehyde is indicated at the top of each gel. The lane labeled "L" is the Fisher BioReagents EZ-run *Rec* protein ladder, and the marker sizes are shown on the left of the gel. The identity of the protein subunits and crosslinked products are shown on the right of the gel. The positions of the dominate band on the gel for each condition are consistent with heterodimer of MSH2-MSH6 (MutSα) or MLH1-PMS2 (MutLα) (2, 3). A faint band with slower migration is seen for MutSα (but not MutLα) and may represent a dimer of MutSα. These results confirm that the crosslinking conditions used in our experiments do not promote non-specific higher-order oligomers of MutSα or MutLα. Finally, the observation that the crosslinked MutSα runs primarily as a heterodimer of MSH2-MSH6 on the SDS gel (C, lane 2) further supports the interpretation of the single peak in the AFM volume distribution (A) being a heterodimer of MSH2-MSH6 (a single MutSα).

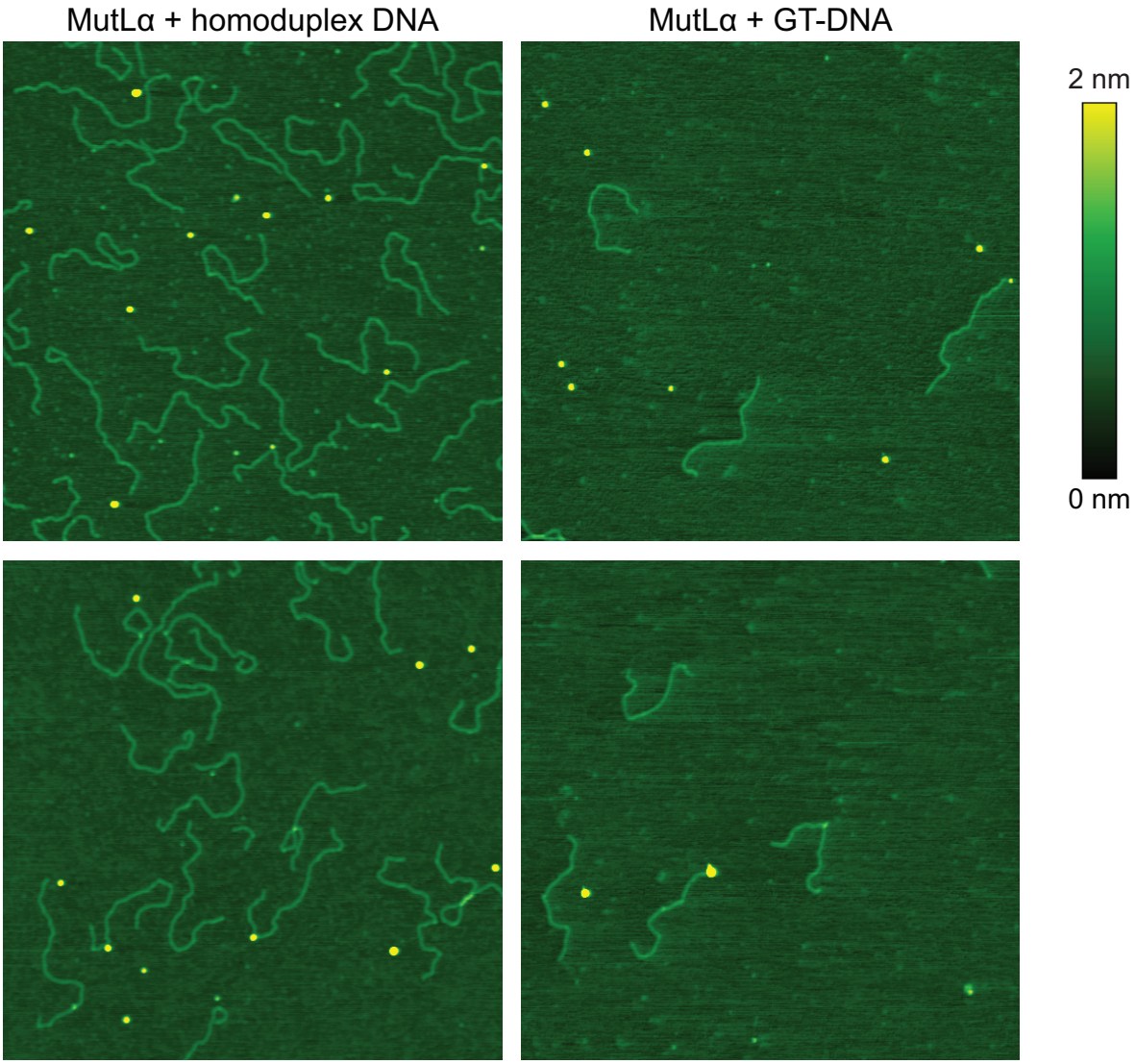

Figure S4. Example images from depositions of DNA crosslinked in the presence of 125 nM MutLα. MutLα, DNA and 1mM ATP were incubated for 2 or 5 minutes, crosslinked for 1 minute, diluted 10-fold and deposited on APTES-treated mica (Methods). MutLα + homoduplex DNA (pUC19-VSR, 2.7 kbp) incubated for 2 minutes (Left column) and MutLα + GT-DNA (2.0 kbp) incubated for 5 minutes (Right column) prior to crosslinking and deposition. Under all conditions less that 3% of the DNA are bound by protein (n>150). Two images are shown for each condition: one (lower) specifically chosen to show MutLα bound to DNA.

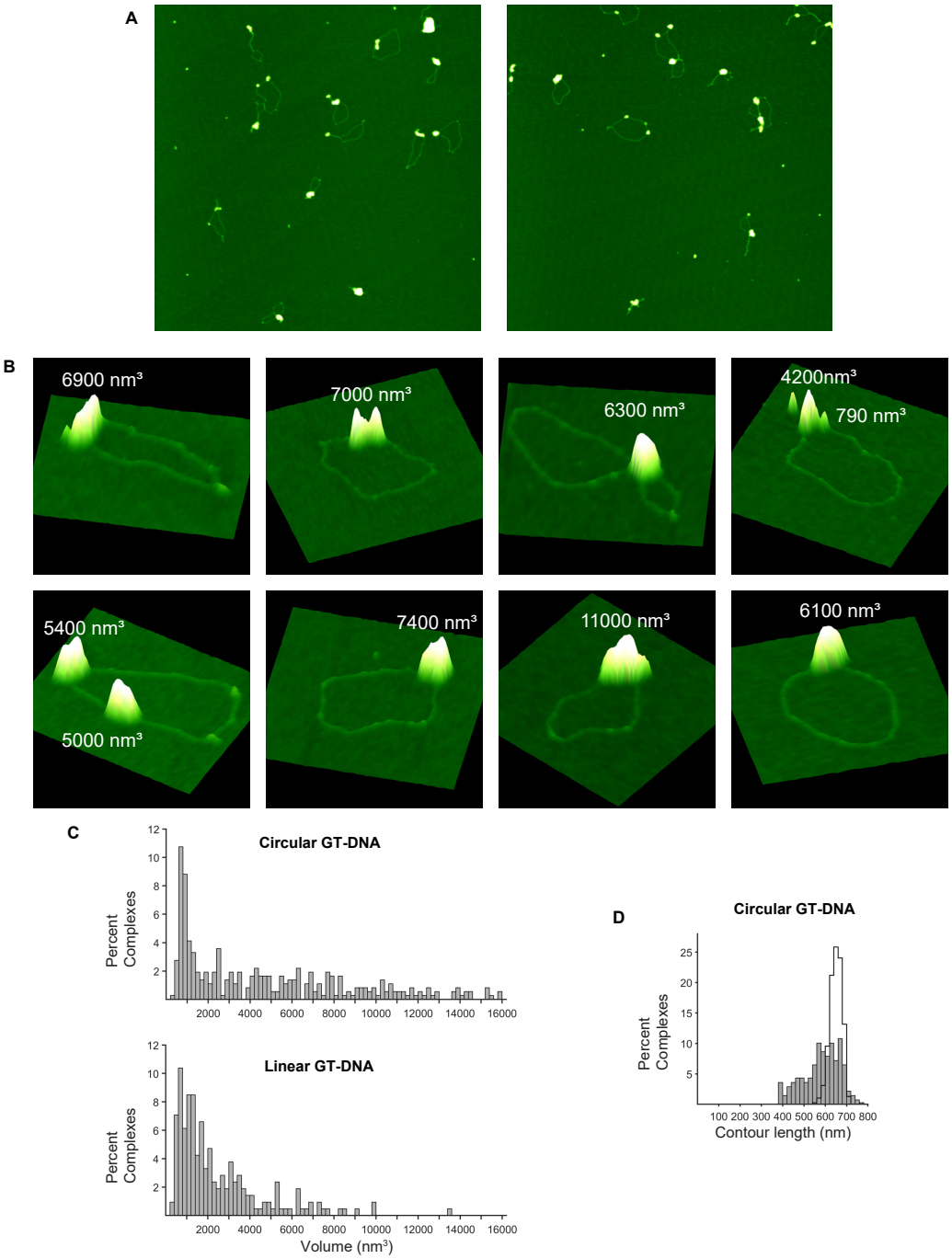

Figure S5. AFM images and analysis of protein-DNA complexes formed on nicked plasmid GT-DNA in the presence of MutSα, MutLα and ATP. (A) Representative 2 μm x 2 μm top-view images showing complexes formed on nicked plasmid (circular) GT-DNA incubated in the presence of 125 nM MutSα, 125 nM MutLα and 1 mM ATP for 2 min prior to crosslinking. **(B)** Examples showing 3-D topographic views of MutSα-MutLα (or MutSα) complexes on circular GT-DNA. The volume of the complex is shown next to it. **(C)** Distribution of volumes for of protein-DNA complexes formed by incubating 125 nM MutSα and 125 nM MutLα with 1 mM ATP and circular (Top) or linear (Bottom) GT-DNA for 2 min prior to crosslinking. The data for linear GT-DNA are the same as those plotted in Figure 3C, except that the x-axis has been extended to 16,000 nm^3. **(D)** Distribution of the contour lengths of circular GT-DNA with protein-DNA complexes (Grey bars) and for free circular GT-DNA (Cityscape). Contour length data for linear GT-DNA are in Figure 3D.

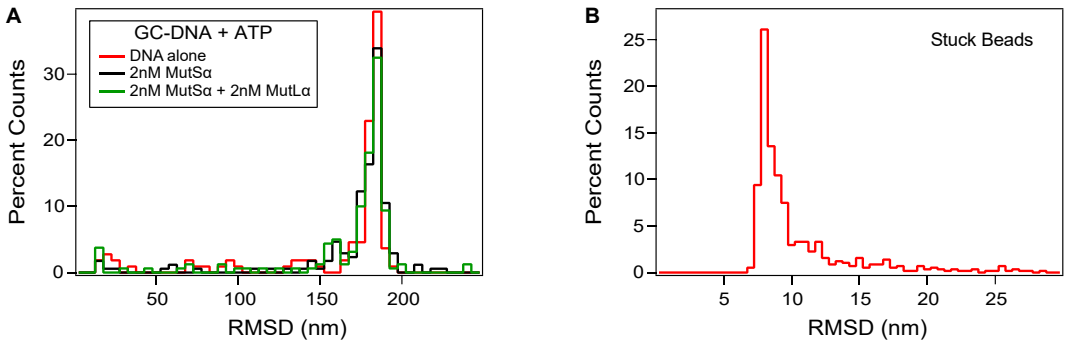

Figure S6. Additional tethered particle motion experiments. Histograms of the root mean square displacements (RMSD) of many DNA molecules using the tethered particle motion assay. **(A)** Experiments with GC-DNA (homoduplex) and 2 mM ATP without protein (red: n= 109), with 2 nM MutSα (black: n=171), and with 2 nM each of MutSα + MutLα (green: n= 161). **(B)** Characterization of lower RMSD limit by analysis of beads stuck to the surface (n=575). Note that the horizontal axis is more expanded in panel B than in panel A.

GT-DNA	One MutSa	Two MutSa	Three or more MutSa
Specific 100uM ADP (n=191)	80%	15%	5%
Non-Specific 100uM ADP (n=26)	88%	12%	0%
Specific 100uM ATP (n=185)	25%	34%	41%
Non-Specific 100uM ATP (n=64)	42%	27%	31%
Specific 1mM ATP (n=72)	52%	25%	22%
Non-Specific 1mM ATP (n=31)	74%	19%	6%
GC-DNA			
100uM ATP (n=353)	78%	14%	8%
100uM ADP (n=156)	69%	15%	13%

Table S1. Percent of MutSα-DNA complexes containing one (volumes: ~600-1200 nm^3), two (volumes: ~1200-2000 nm^3), or three (or more) MutSα proteins for both specific and non-specific complexes. Data shown for different nucleotide conditions with and without a mismatch in the DNA. All data were acquired with 125 nM MutSα.

Supporting Methods

Imaging and Image Analysis

The images were captured in air with a Nanoscope IIIa (Digital Instruments, Santa Barbara, CA), an Asylum Research MFP-3D, or a JPK NanoWizard 4 microscope in tapping mode. Pointprobe Plus tapping mode silicon cantilevers (Nanosensors, Switzerland) with resonance frequencies from 146-236 kHz were used. The images were collected at a speed of 1.97 Hz, a size of 2 μm x 2 μm, and at a resolution of 512 x 512 pixels. DREEM imaging was conducted as described previously (4) using an Asylum MFP3D AFM. A combination of NIH ImageJ64 (Rasbrand, with NeuronJ plug-in) software, Nanoscope III v5.3 software (Veeco, Santa Barbara, CA), ImageSXM v1.95, as described previously (5-7), and a custom MATLAB program (ImageMetrics; available at https://imetrics.app/) were used to measure the volumes of the complexes on the DNA, the DNA contour lengths, and the position of the proteins on the DNA. All methods produced equivalent results. The KaleidaGraph program (Synergy Software, Reading, PA) and MATLAB were used to generate statistical plots for each data set. For each data set, 20-70 images from two to three independent experiments were analyzed, compared, and pooled.

We found that the crosslinking significantly increased the heights of the MutSα and MutLα on the surface, and therefore, their volumes are larger than they would be without crosslinking (See SI Appendix, fig. S3) and do not fit on the standard curve (7). The volume distribution for crosslinked free MutS in the presence of ATP exhibits a single peak centered at ~800 nm^3 (See SI Appendix, fig. S3A), which is similar to MutSα-GT-DNA complexes in the presence of ADP and MutSα-GC-DNA complexes in the presence of ATP (See SI Appendix, fig. S2B, S2C). Consequently, we estimate the numbers of proteins in the protein-DNA complexes based on this volume (~800 nm^3). Notably, the first peak in the ATP experiments (Figure 3A) overlaps with this volume. The number of proteins indicated in the text and figures for SL complexes are rough estimates, based on the volume of MutSα alone. Extrapolation of these estimates to the larger complexes with "missing" DNA is particularly difficult due to contributions from the DNA and the effect that shape and height have on the apparent volumes measured from AFM images (1) (See SI Appendix, legend fig. S3B).

To determine the positions of MutSα binding on the DNA fragments, the distance from the center of the bound MutSα complex to each end of the DNA fragment was measured. Because the mismatch is 124 nm from end of the DNA, there will be a "short arm" and a "long arm" DNA length when MutSα is bound at the mismatch. Complexes with centers ±1 standard deviation of the expected

mismatch position are categorized as specific complexes. We did not end label the DNA to identify the DNA ends; thus, some nonspecific complexes will be counted as specific complexes but not vice versa.

Because the SL complexes can be large and result in shorter DNA contour lengths than the error of the measurements, the positions of the complexes along the DNA were measured at the start of the complex, the end of the complex, and the total length of the DNA was also recorded. The DNA length missing from each DNA molecule was then calculated as the difference between the expected DNA length and the apparent length. We assume the missing DNA length is included in the protein complex, so the missing DNA length is added to the overall protein complex length if the total length is shorter than 1 standard deviation of the measured free DNA length, as shown in Figure 4A. We created displays of the DNA contour lengths and positions for each DNA molecule as shown in Figure 4. This approach avoids bias in measuring the DNA contour length when a large protein complex is present, so that we can account for the DNA length inside the protein complex more systematically. A protein-DNA complex is considered specific if any part of the complex (including the missing DNA length) is within ±1 standard deviation (as determined in the MutSα experiments conducted with ADP) of the mismatch.

Tethered particle motion assay

Tethered particle motion (TPM) experiments measure changes in DNA configurations that result in changes in the DNA end-to-end distances, such as protein-induced DNA bending, DNA looping, or wrapping, by monitoring the Brownian motion of a bead attached to the end of a surface-tethered DNA fragment (8-10). Our TPM experiments were performed in chambers with surfaces passivated by PEG. Holes drilled in glass microscope slides (VWR) allowed buffer exchange in the chambers. The slides and No. 1.5 coverslips (VWR) were cleaned with sequential sonication in acetone, ethanol, potassium hydroxide (1M) and water. The slides were dried in air before adding PEG solutions. Methoxy-poly (Ethylene Glycol)-silane (mPEG-silane, Avg. MW 2,000) and Biotin-Poly(Ethylene Glycol)-silane (Biotin-PEG-silane, MW 3,400) were purchased from Laysan Bio., Inc. 20 mg of mPEG-silane was dissolved in 80 microliters 0.1M sodium bicarbonate solution (250 mg/ml) and 2 mg of Biotin-mPEG-silane was dissolved in 10 microliters 0.1M sodium bicarbonate solution (200 mg/ml). 1 microliter of the Biotin-mPEG-silane solution was added to 80 microliters of mPEG-silane solution to yield a 1% biotinylated mPEG-silane solution. This solution was briefly centrifuged for a few seconds at 13,000 RPM in a micro-centrifuge to remove air bubbles and then was promptly applied to the channel region of the slide. Finally, a coverslip was placed on the slide so that the biotinylated mPEG-silane liquid could be incubated overnight at room temperature in a humid box to prevent the solution from drying.

The next day, the coverslip and slide were separated, both surfaces were rinsed with water, and then air-dried. A second application of mPEG-silane (without the biotin-mPEG-silane) by the same method improved passivation of the surface. The PEG coated slides were stored dry for up to a week.

The tethers were 550 bp, double-stranded DNA with a biotin on one end, a centrally located GT mismatch, and a digoxin attached to the other end. Methods for construction of this DNA are described elsewhere (11). Controls include DNA lacking the GT mismatch. The beads were SPHERO protein G polystyrene particles with 0.84 μm diameter (Spherotech, catalogue number PGP-08-5, supplied at 1% wt/vol). Beads were washed before use to remove free protein G by 4 cycles of pelleting by centrifugation, removing supernatant solution and resuspending in deionized water. The final resuspension was in wash buffer (20 mM Tris:HCl, pH 7.8, 100 mM NaOAC, 5 mM $MgCl_2$) at the same 1% wt/vol concentration supplied by the manufacturer. For experiments, this sample was diluted 20-fold in wash buffer.

Flow channels were formed between the PEG coated slide and coverslip using double sided tape. All samples washed into the channel were suspended in wash buffer. We first added 0.1 mg/ml streptavidin (Invitrogen S888) to the biotin-PEG coated channel, incubated for 5 minutes, and then rinsed with wash buffer. Second, the biotin/digoxigenin end-labeled 550 bp DNA substrate was added at 10 pM, incubated 5 minutes, and rinsed. Third, anti-digoxigenin was added at 4 μg/ml, incubated 15 minutes, and then the channel was rinsed with wash buffer. Fourth, Protein G coated beads were added at 0.05 % wt/vol, incubated 20 minutes, and then the channel was slowly rinsed with wash buffer. Rinsing too vigorously will remove tethered beads from the surface. Finally, mismatch repair proteins were added to the channel in wash buffer containing ATP or other nucleotides as indicated in the text of the paper. MMR proteins were added 2-5 minutes before observations commenced. Data were recorded from samples for times ranging from 5-15 minutes each.

Data were acquired using a microscope (Olympus IX71) illuminating an area of the surface of the flow cell with attached beads with white light using the lamp in the transmission illumination column of the microscope body and imaging with a 60X water immersion objective (Olympus UPlanApo 60x N.A. 1.2) as well as the extra 1.6x built-in magnification option in the microscope body onto an emCCD camera (Cascade 512b, Roper Scientific). The 16 μm x 16 μm square pixels of this camera correspond to 167 nm x 167 nm area of the flow cell surface. Movies of bead motion for homoduplex DNA were acquired at 100 msec per frame and for GT DNA were acquired at 50 msec per frame.

Data analysis used custom MATLAB codes. Beads were identified in a background subtracted image containing the average of the first 10 frames of the movie by locating pixels with intensity above

a threshold of 7 standard deviations of the intensities of pixels across the entire image that were also separated by more than 12 pixels from other identified spots. Regions centered on an identified bead that were 29 pixels by 29 pixels were then isolated for subpixel peakfinding. The radial symmetry approach (12) was applied to the 29 x 29 pixel image centered on each bead to estimate the sub-pixel location of the center of the bead in every frame of ~1800 frame movies. Slow drift was removed from the sub-pixel particle position time trace for each bead by applying a second order Butterworth filter with a highpass cutoff at 0.15 Hz (9). The average x and y locations over the time trace was calculated (x_{avg} and y_{avg}) (omitting the first 50 and last 100 frames). These locations were used to then calculate the root mean squared displacement as

$$RMS\ displacement = \sqrt{\frac{1}{N_{frames}} \sum_{i=1}^{N_{frames}} \left[(x_i - x_{avg})^2 + (y_i - y_{avg})^2\right]}$$

where x_i and y_i are the coordinates of the center location determined in frame i.

Without any added MMR protein in solution, RMS bead displacement histograms centered at 180 nm using a 550 base pair DNA tether, 0.84 μm diameter beads, and 10 Hz or 20 Hz imaging, which is in close agreement with results reported in Han et al. (9) for similar experiments. Additionally, analysis of beads that were stuck to the surface resulted in RMS bead displacement distributions peaked at 7.7 nm with width of 1 nm (Figure S6B). Stuck beads were excluded from our analysis by requiring accepted beads have RMS displacement greater than 15 nm. We also excluded beads with asymmetric motion above a threshold common in the field (9). We performed shape analysis on scatter plots of bead positions identified in every frame of a movie using the MATLAB command regionprops, which reports axes for elliptical fits to these regions. We excluded any beads where the aspect ratio (MajorAxisLength/MinorAxisLength) was greater than 1.07. Control experiments with homoduplex DNA (lacking any mismatch) showed RMS motion of 180 nm with MutSα + ATP or MutSα + MutLα +ATP in solution, which is the same RMS motion as having no protein in solution (Figure S6A). The statistical significance of the differences between RMSD distributions for different conditions was evaluated using the two sample Kolmogorov-Smirnov test (13). Due to non-specific interactions of the MutSα and MutLα proteins with the surface, these experiments were conducted at lower protein concentrations than the AFM experiments, such that no non-specific interactions were observed with homoduplex DNA. These experiments also differ from the AFM experiments in that the DNA end is blocked, so unlike in the AFM experiments MutSα mobile clamps cannot slide off the end of the DNA.

References

1. G. Varadhan, W. Robinett, D. Erie, R. Taylor, *Fast simulation of atomic-force-microscope imaging of atomic and polygonal surfaces using graphics hardware*, Electronic Imaging (SPIE, 2002), vol. 4665.

2. S. Acharya *et al.*, hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proc Natl Acad Sci U S A* **93**, 13629-13634 (1996).

3. M. C. Hall *et al.*, DNA binding by yeast Mlh1 and Pms1: implications for DNA mismatch repair. *Nucleic Acids Res* **31**, 2025-2034 (2003).

4. D. Wu *et al.*, Visualizing the Path of DNA through Proteins Using DREEM Imaging. *Mol Cell* **61**, 315-323 (2016).

5. G. C. Ratcliff, D. A. Erie, A novel single-molecule study to determine protein--protein association constants. *Journal of the American Chemical Society* **123**, 5632-5635 (2001).

6. H. Wang *et al.*, DNA bending and unbending by MutS govern mismatch recognition and specificity. *Proc Natl Acad Sci U S A* **100**, 14822-14827 (2003).

7. Y. Yang, H. Wang, D. A. Erie, Quantitative characterization of biomolecular assemblies and interactions using atomic force microscopy. *Methods* **29**, 175-187 (2003).

8. B. Henneman, J. Heinsman, J. Battjes, R. T. Dame, Quantitation of DNA-Binding Affinity Using Tethered Particle Motion. *Methods Mol Biol* **1837**, 257-275 (2018).

9. L. Han *et al.*, "Calibration of tethered particle motion experiments" in Mathematics of DNA Structure, Function and Interactions. The IMA Volumes in Mathematics and its Applications, B. C, H. S, O. W, S. D, S. D, Eds. (Springer, New York, NY, 2009), vol. 150, pp. 123-138.

10. D. T. Kovari, Y. Yan, L. Finzi, D. Dunlap, Tethered Particle Motion: An Easy Technique for Probing DNA Topology and Interactions with Transcription Factors. *Methods Mol Biol* **1665**, 317-340 (2018).

11. R. Qiu *et al.*, MutL traps MutS at a DNA mismatch. *Proc Natl Acad Sci U S A* **112**, 10914-10919 (2015).

12. R. Parthasarathy, Rapid, accurate particle tracking by calculation of radial symmetry centers. *Nat Methods* **9**, 724-726 (2012).

13. T. W. Kirkman (1996) Statistics to Use. [Online]. Accessed 06 August 2019: http://www.physics.csbsju.edu/stats/KS-test.html