DNA target sequence identification mechanism for dimer-active protein complexes

Markita P. Landry, Xueqing Zou, Lei Wang, Wai Mun Huang,

Klaus Schulten, Yann R. Chemla

SUPPLEMENTARY METHODS

QD labeling efficiency

TelK QD labeling efficiency was estimated using a gel-shift assay. A non-cutting mutant of full-length TelK was used, TelK-YF, for which the active site tyrosine was mutated to a phenylalanine. TelK-YF binds to target-site DNA as efficiently as wild-type TelK, but does not cut the substrate (26). A 66-bp DNA duplex containing the TelK target site and labeled with a single 5'-Alexa647 dye (Integrated DNA Technologies, Coralville, IA) served as a substrate. 10 ng of Alexa-labeled DNA was incubated with 50 nM unlabeled TelK-YF and separately with 50 nM QD-labeled TelK-YF (labeling technique described in **Materials and Methods**) for 1 hour at room temperature.

The Alexa-DNA substrate by itself, the substrate with unlabeled TelK-YF, and the substrate with QD-labeled TelK-YF were run on a 0.5% agarose gel for 1 hour at 100 V. This gel was imaged using 650-nm excitation/680-nm emission to visualize the Alexa-DNA (**Supplementary Figure S1a)** and using 525-nm excitation/550-nm emission to visualize the QDs (**Supplementary Figure S1b)** . The DNA bound to unlabeled TelK-YF shifted slightly (**Supplementary Figure S1a** , lane 2), and DNA bound to QD-TelK-YF shifted more significantly (lane 3). QD emission was only observed at the DNA + QD-TelK shifted band position (**Supplementary Figure S1b** , lane 3). We calculated the labeling efficiency from the percentage of Alexa-DNA bound by QD-TelK, as determined by the ratio of the DNA + QD-TelK shifted band intensity to the total DNA intensity:

$$\text{Labeling efficiency} = \frac{I_{\text{shifted DNA+QD TelK band}}}{\sum I_{\text{lane}}}$$

This procedure yielded a labeling efficiency of ~84%.

Precautions were taken to prevent QD photobleaching prior to imaging: QDs were stored under dark and only exposed to light once in our TIRF setup, and we avoided reducing agents in our buffers. As a result, we expect there to be only a negligible fraction of dark QDs prior to imaging that could affect our labeling efficiency estimates. We also performed a test of QD stability under our imaging conditions. We illuminated our glass chamber with TelK-QD stuck to the surface of our microfluidic chamber for 500 seconds, and found an average QD lifetime of 277 ± 18 s (mean ± s.e.m.). After a few minutes of TIRF illumination, it became difficult to distinguish between protein dissociation and QD photobleaching. Due to the increased error that could be introduced into our data, we limited our movies to ~1 minute.

QD labeled TelK shows wild-type activity

QD-labeled TelK cutting activity was tested to ensure that modification of TelK via the noncovalent attachment of Anti-His QDs did not affect its function. Wild-type TelK was labeled with QDs as described in **Materials and Methods** . Both QD-labeled and unlabeled 50-nM TelK samples were incubated with 50 ng of 4.3-kb pSKN DNA, synthesized as described in **Materials and Meth ods** to contain the TelK target sequence slightly off-center at position 2092. TelK is a single-turnover protein (25), ensuring that only one active TelK dimer was responsible for cutting each DNA substrate molecule. QD-TelK and pSKN DNA were incubated for 15 minutes at room temperature, and subsequently run on a 1% agarose gel as detailed in (25). The activity was determined from the ratio of the cut DNA band intensity to that of the input DNA

$$\text{Activity} = \frac{I_{\text{cut DNA}}}{\sum I_{\text{lane}}}$$

where *I* is the band intensity at 535 nm measured with a gel scanner (Kodak 4000mm). Both lanes showed complete cutting of DNA by TelK (**Supplementary Figure S1c**). Therefore, the activity of labeled TelK was determined to be unaffected by the QD, as compared to unlabeled TelK.

This assay does not determine directly the rate of target localization for the QD-labeled TelK. Though previous studies using QD-labeled proteins have shown a slight decrease in the protein diffusion coefficient due to QD size, motion of the labeled protein has been shown to remain representative of the unlabeled protein behavior (56).

SUPPLEMENTARY DISCUSSION

Small DNA Condensation Step Size Could be Due to Misoriented TelK Dimers

Our optical trap studies show that TelK dimerizes non-target DNA. Two distinct condensation step sizes are observed from the optical trap data, one of which matches what would be expected from the TelK-DNA crystal structure. However, the second condensation step size (which accounts for 51% of all condensation events) remains puzzling. We propose that the small step size is due to the interaction of two misoriented TelK monomers on DNA.

Previous work has proposed that the high degree of cooperative electrostatic binding between protein monomers in dimer- or oligomer-active protein families may lead to different modes of nonspecific binding that depend on protein orientation on the DNA. For example, in the process of site-specific recombination, the two different modes of binding between two recombinase monomers determine which DNA strand is cut first, and determine the product of recombination (57). Correct orientation of two monomers in dimer-active Y-recombinases has also been proposed to be a control mechanism that ensures correct protein-DNA interactions prior to catalysis (58,59). As a member of the Y-recombinase protein family, TelK must correctly orient itself along the DNA target site in order to form a functional unit. The TelK crystal structure shows that the active dimer must be oriented head-to-

head on the dyad symmetric DNA target site, and shows that this protein-DNA interaction causes a sharp 73° bend in the target DNA (26). However, due to the random orientation of TelK monomers along DNA, we speculate that half of the stochastic monomer encounters along DNA will involve dimers that are incorrectly oriented in a head-to-tail or tail-to-tail orientation. These dimerization events may explain the bimodal distribution of condensation step size, with the large DNA condensation corresponding to correctly oriented TelK monomers, and the small DNA condensation corresponding to incorrectly oriented molecules.

When repeating the optical trap measurements with target DNA, we observed similar condensation events that yielded an average condensation step size of 7.75 nm. In contrast to non-target DNA, the step size distribution for target DNA exhibited a single peak, and we did not observe as many condensation events at ~4.5 nm (**Supplementary Figure S6**). We speculate that the presence of the target-site may increase the likelihood of observing correctly oriented TelK dimers. Our TIRF assay shows that TelK monomers have target-site specificity and our stochastic simulations show that target finding occurs approximately an order of magnitude faster than dimerization. Therefore, it is likely that monomers immobilize at the target site with the correct orientation along the DNA axis, thus increasing the probability of forming a correctly oriented dimer.

Absence of TelK hairpin formation on linearly extended DNA substrates

When using a DNA substrate containing the TelK target sequence, it was expected that TelK would cut the DNA at the target site and form DNA hairpins. However, no DNA cutting or hairpin formation was observed with DNA containing the TelK target site in either the TIRFM or optical trap assay in over 10 minutes. In both assays, the activity of TelK is observed on DNA substrates that are linearly extended, either in the form of DNA bridges in TIRFM or DNA tethers in the optical trap. These observations strongly suggest that TelK activity may be inhibited on linear DNA substrates extended by

applied low tensions. To confirm this hypothesis, we performed a tethered particle microscopy (TPM) assay which allowed us to test the single-molecule activity of TelK on our DNA substrates in a globular instead of a linearly extended conformation. This assay involved tethering one end of a DNA substrate to a glass surface and tethering the other end to a fluorescent microsphere. Here, we observed that TelK cut a 4.5-kb globular target DNA substrate at a rate of one cut per 41.6 ± 71.4 s (mean ± s.e.m.). In essence, while TelK is able to search for, find, and preferentially bind to its target site on linearly extended DNA substrates, it is unable to catalyze the formation of DNA hairpins on extended DNA. We propose that the process of DNA cutting and hairpin formation is very sensitive to DNA conformation, and that even applications of low force (3.1 ± 0.2 pN for TIRFM DNA, 5.2 ± 0.1 pN optical trap DNA; mean ± s.e.m.) inhibit TelK activity. Future studies will address the process of TelK hairpin formation in greater detail.

Simulations predict a power-law dependence of protein activity on protein concentration

Our dimer-active model predicts a power-law dependence of protein activity on protein concentration. Stochastic simulations of two kinetic processes—dimerization and target localization—over a large range of protein concentrations show a difference in the power-law dependence for each of these processes. This can be understood from a simple scaling argument, considering a finite length of DNA coated with N protein monomers when the process of target-finding and dimerization are both diffusion limited (and not protein binding-limited). On average, the distance between the protein closest to the target site is inversely proportional to the number of proteins on the DNA ($\langle d_{min} \rangle \sim 1/N$). Conversely, the average shortest inter-protein distance varies inversely with the square of the number of proteins ($\langle x_{min} \rangle \sim 1/N^2$) . Thus, the rates for target localization and for dimerization are given by the protein diffusion rate over these distances, and will depend differently on protein occupancy:

$$k_{\text{target}} \sim {}^{D}/_{\langle d_{min} \rangle^2} \sim N^2$$

and

$$k_{\text{dimer}} \sim {}^{D}/_{\langle x_{min} \rangle^2} \sim N^4.$$

These relationships are observed in our simulations (**Figure 6b**, blue and red points) under conditions where the rate-limiting step was protein diffusion on DNA: k_{dimer} (~$N^{4.4}$; $R^2 = 0.98$), and k_{target} (~$N^{2.0}$; $R^2 = 0.97$). At low occupancies, where binding of protein is expected to be the rate-limiting step, a different scaling is expected (**Figure 6b**, black points): k_{dimer} (~$N^{2.2}$; $R^2 = 0.99$).

SUPPLEMENTARY TABLE

Name	Protein	DNA	Number of atoms	Time (ns)
SimuA	TelK dimer	Target DNA	347,912	80
SimuB	TelK dimer	Non-target DNA	343,306	80
SimuC	TelK monomer	Target DNA	281,097	80
SimuD	TelK monomer	Non-target DNA	276,057	80
SimuE	TelK dimer	None	280,741	80

Table S1: List of performed MD simulations

SUPPLEMENTARY FIGURES

Figure S1.

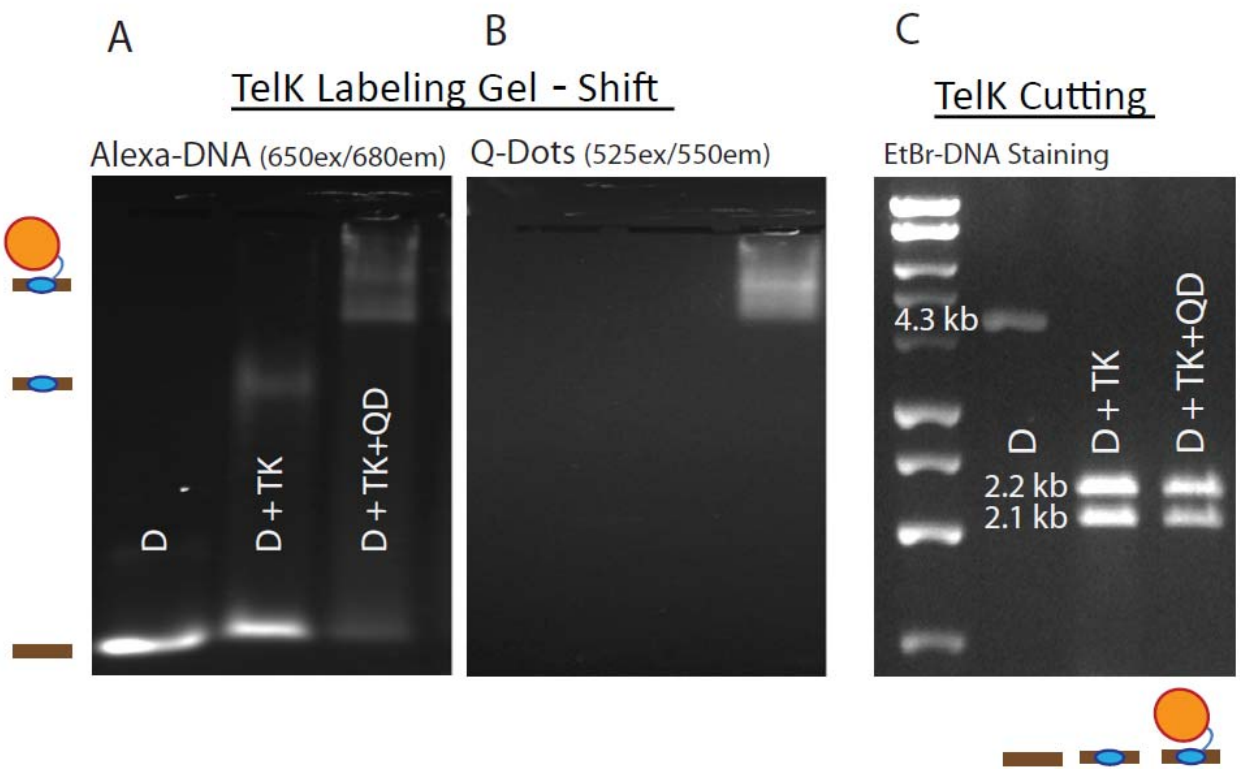

Figure S1. Quantification of TelK Quantum Dot labeling efficiency, and labeled TelK activity.

(a) Gel shift quantification of TelK-bound DNA. Gel imaged to visualize Alexa-DNA. Lanes, from left to right: Alexa-DNA only, Alexa-DNA with TelK, and Alexa-DNA with QD-labeled TelK. (b) Same gel imaged to visualize QDs. Intensity analysis shows an 84% labeling efficiency. (c) Quantification of TelK cutting activity. Lanes, from left to right: 1-kb DNA ladder, pSKN (target) DNA only, with TelK, and with QD-labeled TelK. Both TelK and QD-labeled TelK cut DNA substrate with the same efficiency, as determined by the DNA band intensities (**Materials and Methods**).

Figure S2.

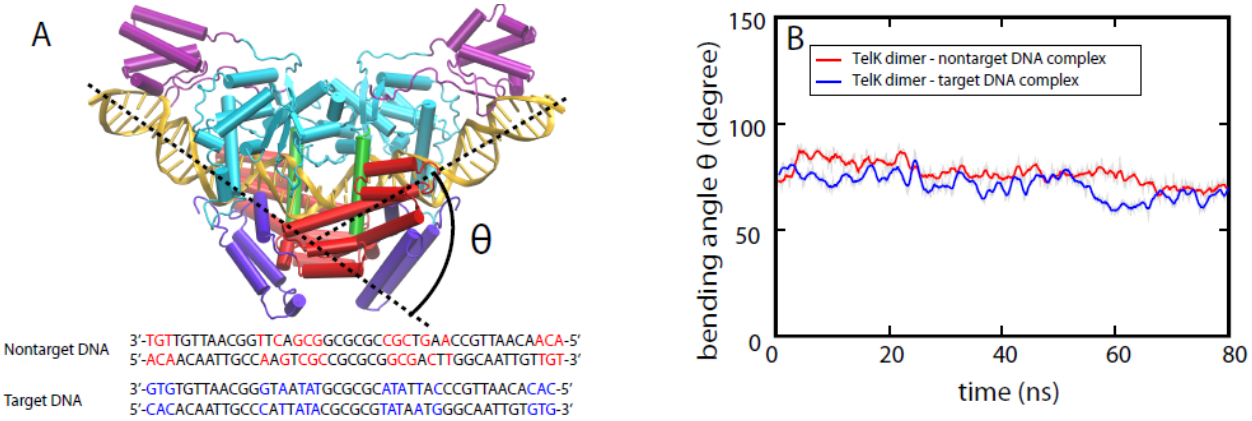

Figure S2. TelK dimer bends target and nontarget DNA.

(a) MD simulation of TelK dimer complexed with non-target (red) and target (blue) DNA. Mutations made to target DNA to produce non-target DNA are highlighted in red (mutated base) and blue (wild-type base). (b) Non-target (red) and target (blue) DNA is bent by similar angles during an 80-ns simulation.

Figure S3.

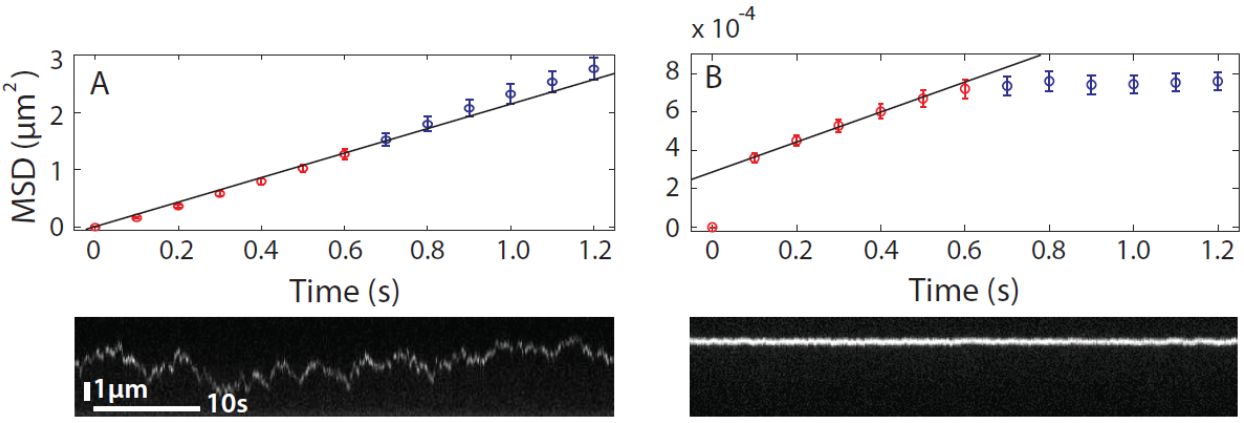

Figure S3. Determination of QD-labeled TelK diffusion coefficients

Representative trajectories and corresponding MSD vs. time plots for mobile (a) and immobile (b) spots are shown. Diffusion coefficients are determined from fitting the first 0.6 s (red data points) of the MSD to a line (**Materials and Methods**). Fits yield diffusion coefficients of D_{mobile} = 1.1 μm^2/s and $D_{immobile}$ = 4.6 × 10^{-4} μm^2/s, respectively. Error bars represent s.e.m.

Figure S4.

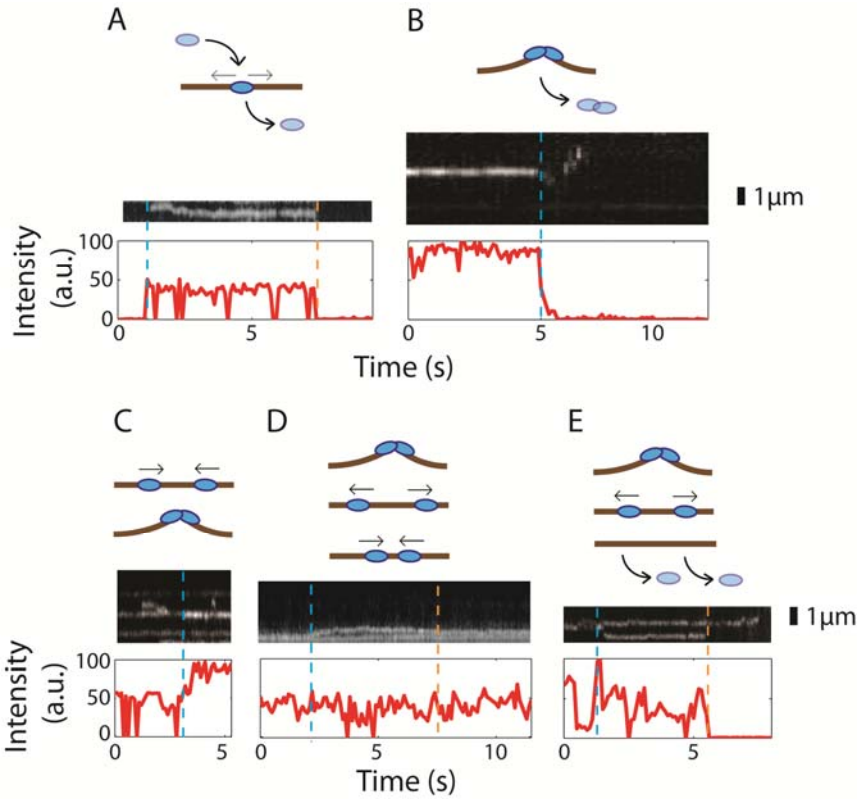

Figure S4. Representative data traces of TelK behavior on non-target DNA.

Schematics of various TelK behavior with corresponding kymographs and fluorescence intensity are shown. (a) A TelK monomer binds to DNA (blue dashed line), diffuses, and dissociates (orange dashed line). (b) An immobile spot dissociates from the DNA (blue dashed line). (c) Two mobile TelK coalesce on DNA as intensity doubles (blue dashed line). (d) An immobile spot separates into two mobile spots (blue dashed line) and re-joins downstream (orange dashed line). (e) An immobile spot separates into two spots (blue dashed line), one of which dissociates from the DNA (orange dashed line). In c-e, the intensity of only one fluorescent spot is shown.

Figure S5.

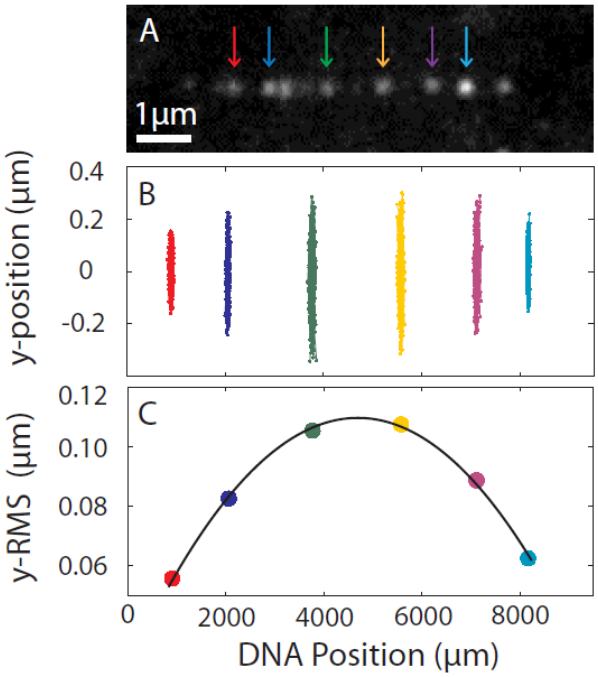

Figure S5. Determination of tether tension in TIRFM assay.

(a) TIRFM image of six immobile spots on a DNA bridge (colored arrows). Tensions experienced by the DNA bridges in the TIRFM assay were estimated from the transverse fluctuations of the DNA by monitoring the y-fluctuations of highlighted spots. (b) The x-and y-position trajectories of the six immobile spots are shown, each of a different color. These thermally driven fluctuations are related to the force at which the bridge is extended by the relation $\langle y_{DNA}^2 \rangle = \frac{k_B T}{F} \sqrt{\frac{l(l_{tot}-l)}{l_{tot}}}$, where y_{DNA} is the motion of the protein spot perpendicular to the DNA bridge axis, k_B is Boltzmann's constant, T is room temperature, F is the force, l is the position of the protein spot, and l_{tot} is the total length of the DNA bridge. (c) Corresponding fit of the variance in transverse Brownian fluctuations $\langle y_{DNA}^2 \rangle$ used to determine the tension on the DNA tether. The transverse positions from individual stationary QD-TelK were analyzed as described in **Materials and Methods** and fit to the above formula using as fitting parameters F, l_{tot}, and a position offset along the glass pedestals (black line). Fits yielded an average tension of 3.1 ± 0.2 pN (mean ± s.e.m.) for all DNA bridges.

Figure S6.

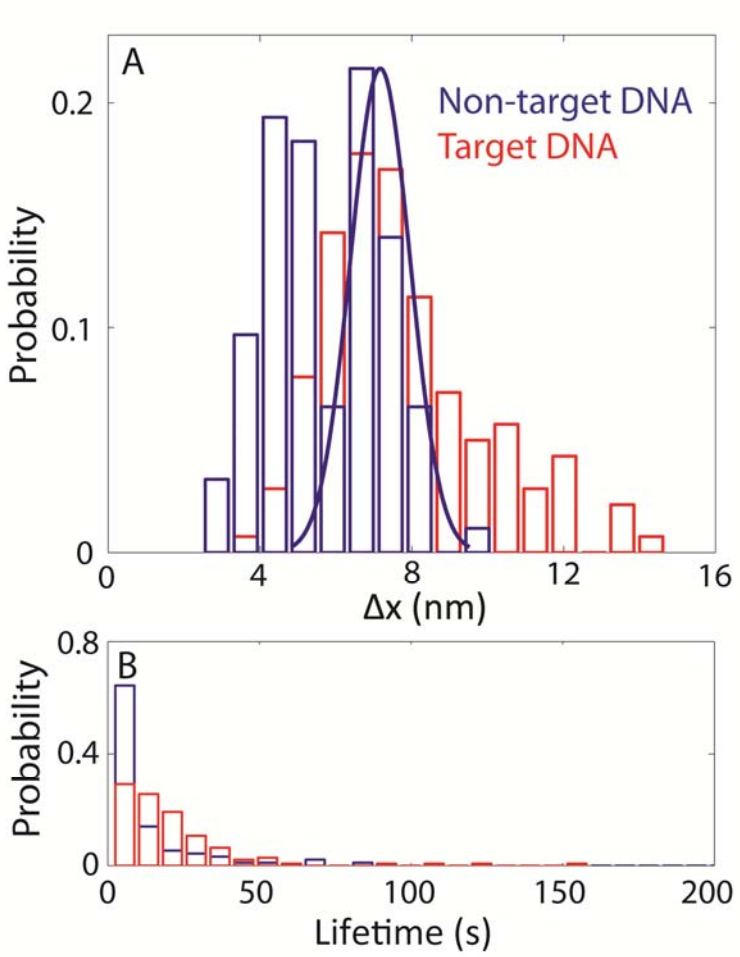

Figure S6. Condensation behavior of target vs. non-target DNA. (a) Step-size distributions for target (red) and non-target (blue) DNA yield a single peaked distribution with an average condensation step-size of 7.75 ± 0.2 nm, and a doubly-peaked distribution with peak condensation step-sizes of 7.2 ± 0.1 nm and 4.5 ± 0.1 nm, respectively. (b) Mean condensed state lifetime distributions for target DNA (red) and for non-target DNA (blue). Mean lifetimes are 22 ± 6 s and 21 ± 2 s, respectively (mean ± s.e.m.).

Figure S7.

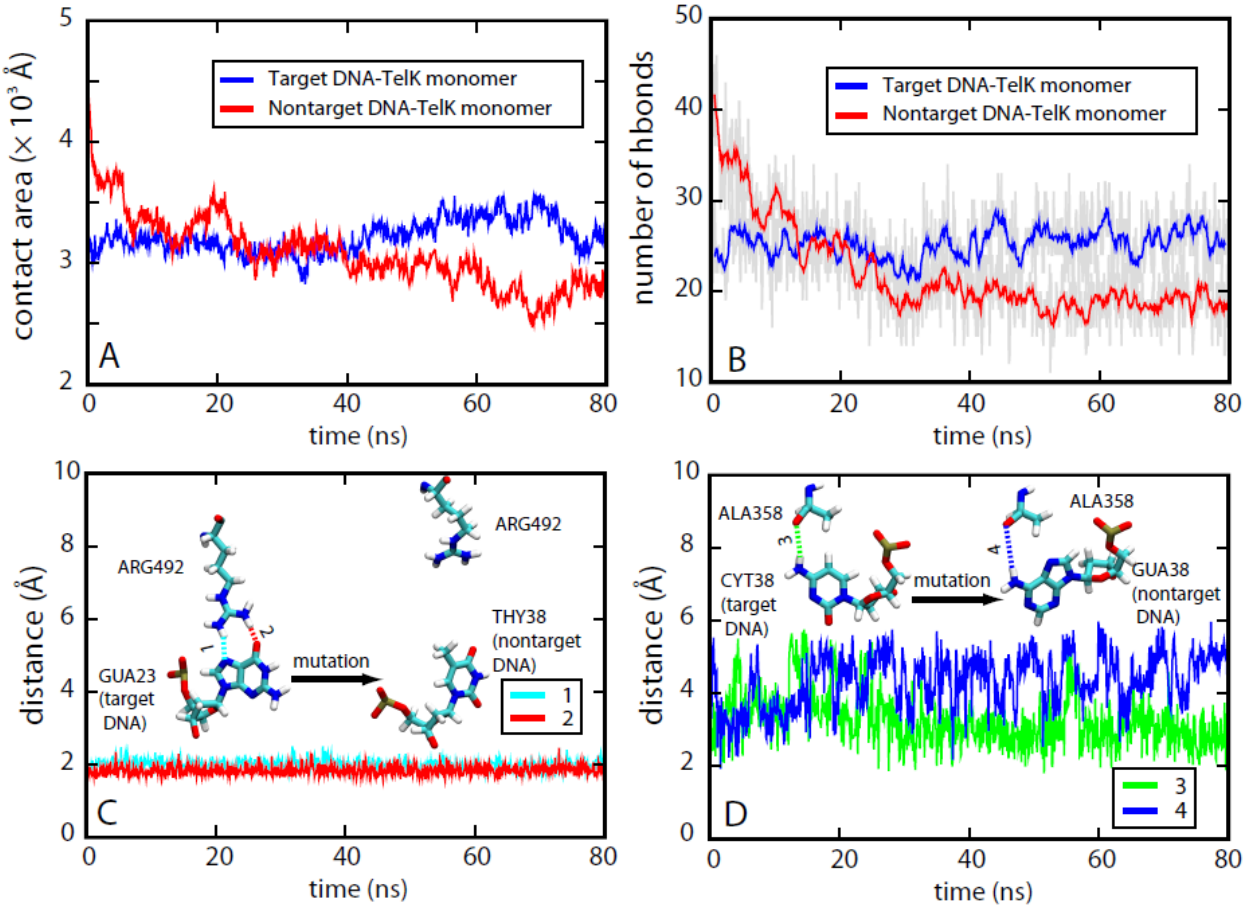

Figure S7. Interaction between TelK monomer and target/nontarget DNA.

(a) Comparison of the contact area between TelK monomer and target DNA (blue) and that between TelK monomer and non-target DNA (red) in MD simulations. (b) Comparison of the number of hydrogen bonds between TelK monomer and target DNA (blue) and that between TelK monomer and non-target DNA (red). Data from (b) and (c) show that TelK monomer binding to target DNA is more stable than binding to non-target DNA; there is an increase in contact area and hydrogen bonds between TelK monomer and target DNA compared to between TelK monomer and non-target DNA. (c) and (d) show two examples of how target DNA mutation affects TelK-DNA interaction. DNA mutations were made based on residues within the TelK footprint as determined in previous studies (25). (c) GUA23 forms two stable hydrogen bonds with ARG492. After mutating GUA23 to a thymine, these two

hydrogen bonds between DNA and TelK are lost. (d) CYT38 forms one stable hydrogen bond with ALA358. After mutating CYT38 to guanine, although a hydrogen bond was observed between GUA38 and ALA358 at the beginning of the simulation, the distance between acceptor and donor atom increased gradually, indicating the instability of that hydrogen bond.

Figure S8.

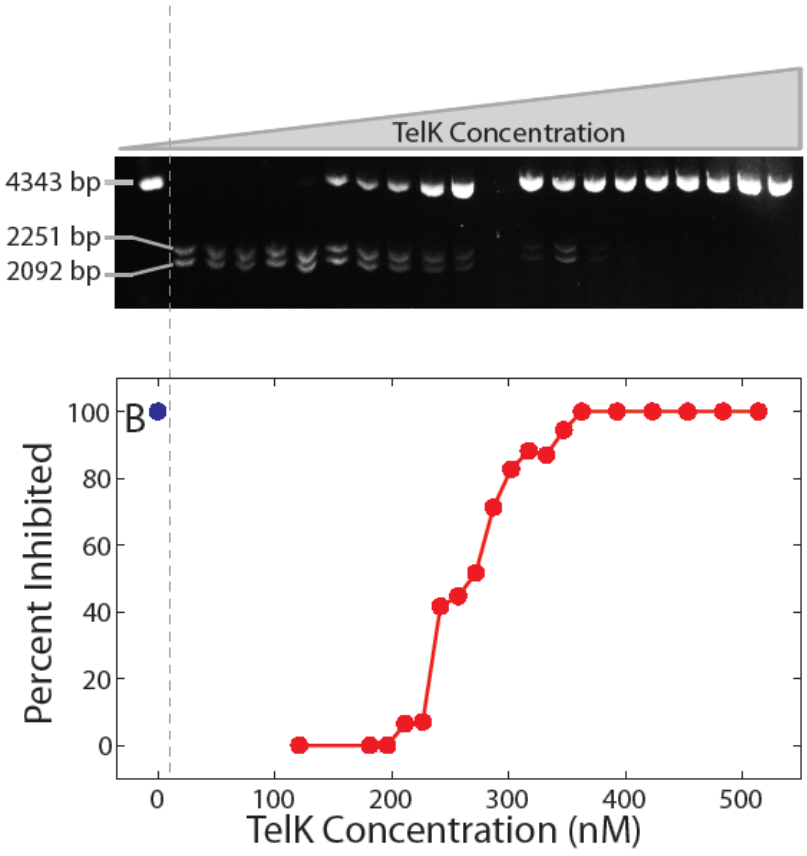

Figure S8. TelK cutting inhibition increases with TelK concentration.

(a) 50 ng of pSKN DNA substrate containing the TelK target sequence slightly off-center was synthesized as described in **Materials and Methods.** This DNA was incubated with 0 nM TelK (DNA-only control) and a range between 120 nM to 513 nM TelK visualized on an agarose gel. DNA cut by TelK is observed as two products at 2092 and 2251 bp. (b) Intensities of each band were measured and used to determine percent inhibition of DNA cutting by TelK via the following relation: $Inhibition = \frac{I_{uncut}}{\sum I_{band}}$.

Figure S9.

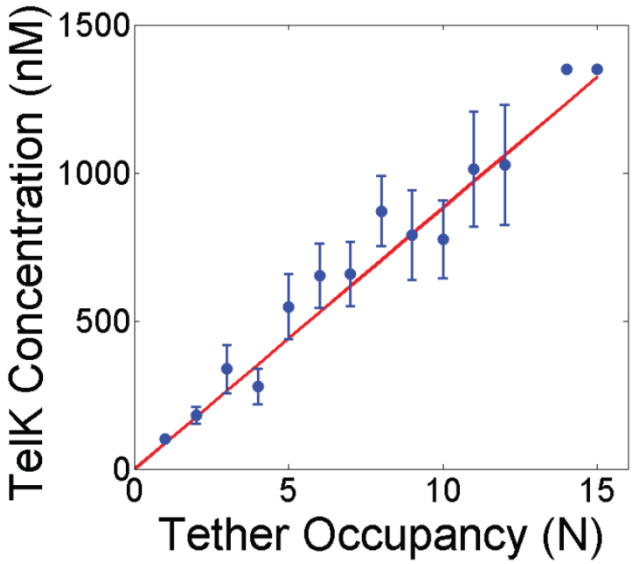

Figure S9. Determination of TelK occupancy on DNA as a function of TelK concentration

From our TIRFM assay, the average number of fluorescent spots visualized on a single DNA bridge is linearly dependent on the TelK concentration at which it was incubated (blue, mean ± s.e.m.; with linear fit in red). To obtain the best estimate of protein number (the TelK occupancy) based on fluorescent spot counts, we used an occupation correction factor that counts non-blinking TelK spots as 2 monomers instead of 1 ($N_{\mathrm{mon}} = 1$; $N_{\mathrm{dim}} = 2$). To extrapolate from the occupancy observed in our TIRFM assay to that expected in our optical trap assay, we accounted for the DNA length difference between TIRFM λ-DNA and optical trap (OT) assay substrates with a length conversion factor, $\xi = \frac{l_{\mathrm{OT}}}{l_{\lambda}}$. This approximation is valid based on the observed uniform binding affinity by TelK on non-target DNA. Thus, for each TelK concentration, the occupancy for optical trap measurements was estimated from the observed TIRFM occupancies by the following relation: $N_{\mathrm{OT}} = N_{\mathrm{TIRFM}}\, N_{\mathrm{mon/dim}} \xi$.

SUPPLEMENTARY REFERENCES

56. Biebricher, A., Wende, W., Escude, C., Pingoud, A. and Desbiolles, P. (2009) Tracking of Single Quantum Dot Labeled EcoRV Sliding along DNA Manipulated by Double Optical Tweezers. *Biophysical Journal*, 96, L50-L52.

57. Grindley, N.D., Whiteson, K.L. and Rice, P.A. (2006) Mechanisms of site-specific recombination. *Annu Rev Biochem*, 75, 567-605.

58. Grainge, I. and Jayaram, M. (1999) The integrase family of recombinase: organization and function of the active site. *Mol Microbiol*, 33, 449-456.

59. Lee, L. and Sadowski, P.D. (2005) Strand selection by the tyrosine recombinases. *Prog Nucleic Acid Res Mol Biol*, 80, 1-42.