Supporting Information for:

Direct Observation of the Formation of CRISPR-Cas12a R-loop Complex at the Single-Molecule Level

Yang Cui[a,b], Yangchao Tang[c], Meng Liang[a,b], Qinghua Ji[a], Yan Zeng[a], Hui Chen[a], Jie Lan[a,b], Peng Jin[c], Lei Wang[c], Guangtao Song[a,b,*], and Jizhong Lou[a,b,*]

[a] Key Laboratory of RNA Biology, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

[b] University of Chinese Academy of Sciences, Beijing 100049, China

[c] Center of Ultra-Precision Optoelectronic Engineering, Harbin Institute of Technology, D-401 Science Park, No. 2 Yikuang Street, Harbin 150080, China

*Email: jlou@ibp.ac.cn

*Email: gsong@moon.ibp.ac.cn

Experimental methods

Recombinant AsCas12a Expression and Purification

AsCas12a and all other mutants was cloned into a custom pET-based expression vector encoding an N-terminal GST-tag followed by His6-tag and a Thrombin protease cleavage site. Point mutations were introduced into AsCas12a using site-directed mutagenesis and verified by DNA sequencing. Cas12a was expressed in Escherichia coli Rosetta 2 (DE3) (Novagen) and purified by chromatography on Ni-NTA Superflow (QIAGEN). After removing the tags by thrombin at 4 °C overnight, cation exchange (SP sepharose) chromatography were used for further purification steps. The final products were evaluated by 10% SDS-polyacrylamide gel electrophoresis (Figure S1).

In vitro transcription and purification of RNA

crRNAs were in vitro transcribed using T7 in vitro transcription kit (MEGAscript T7, Invitrogen) and PCR generated DNA templates carrying a T7 promoter sequence. RNA product was gel-purified before use.

DNA cleavage assay

pUC19 based protospacer plasmids for in vitro cleavage assays were generated by annealed oligonulceotides between digested EcoR I and Hind III sites in PUC19. Ligated plasmids DNA were transformed into E. Coli DH5a cells according to a standard heat shock protocol. Plasmid cleavage assays were carried out by using a custom-designed plasmid containing a 24-bp DNA target sequence and a 5'-TAAA-3' PAM motif. Purified AsCas12a (final concentration 125 nM) and synthesized sgRNA (final concentration 375 nM) were pre-incubated in the cleavage buffer (20 mM HEPES, 100 mM KCl, 2 mM $MgCl_2$) at 37 °C for 15 min, 500 ng supercoiled plasmid were then added to the reaction mixtures and incubated at room temperature for various time points. The reactions were stopped with 6X SDS loading buffer prior to loading on the agarose gel. Cleavage products were run on 1.2% agarose gels (1×TAE buffer) and Midori Green (NIPPON Genetics Europe) staining.

Electrophoretic mobility shift assays

Alexa-488 labeled DNA duplex were prepared by PCR amplification of above plasmid DNA target with 5'-end Alexa-488 labeled primers. Binding assays were performed in buffer containing 20mM HEPES pH 7.5, 100mM

KCl, 2 mM $MgCl_2$, in a total volume of 20 µl. dCas12a mutant was programmed with equimolar amounts of crRNA and titrated from 100 pM to 1 µM. Alexa488-labeled DNA was added to a final concentration of 0.1 µM. Samples were incubated for 1 h at 37°C and resolved at 4°C on an 15% native polyacrylamide gel containing 1X TBE and 2mM $MgCl_2$. DNA was visualized by phosphorimaging, quantified with ImageQuant (GE Healthcare).

FRET-based DNA binding assay

DNA duplex with Cy3 labeled at the 5' of the target strand, and BHQ2 labeled at the 3' of the non-target strand (Figure S10A) was purchased from Invitrogen (Thermo Fisher Scientific, Shanghai). BHQ2 quenches the emission of Cy3 when they are close to each other, thus CY3 fluorescence can only be detected after the disruption of the duplex (for example, after R-loop formation by Cas12a). The dual-labeled DNA duplex was first annealed and diluted into 50 nM aliquots. dCas12a and crRNAs were pre-incubated at 37°C for 10 min in reaction buffer before being incubated with the target DNA. The Cas12a-crRNA with different concentrations was then added into the DNA solution and incubated for ~30 min at room temperature. Fluorescence measurements were collected with a 5 mm path-length quartz cuvette (Hellma Analytics) and a F-7000 (Hitachi), using 5 nm slit widths and 0.2 s integration time. The samples were excited at 532 nm and emitted light was collected from 550 – 700 nm. Data processing and binding fits were conducted using Prism software (GraphPad Software).

Preparation of hairpin DNA constructs.

The DNA hairpin constructs used in our experiments were prepared using the protocol made by Woodside et al. [29] Briefly, two dsDNA handles (2000 bp) was prepared by means of Autosticky PCR using Lambda-phage DNA (New England Biolabs N3011) as the template, one 5' digoxigenin- or biotin-labeled primer (Sangon Biotech), and one primer with a 15-nt extension after a deoxyribose spacer site (Sangon Biotech) to achieve the specified dsDNA length and 5'-end. The hairpin construct was prepared through the ligation of two oligonucleotides to the two DNA handles as in Figure S2. The final product was further gel-purified before use. These constructs were incubated with 2-µm diameter polystyrene beads labeled with streptavidin and 3-µm diameter beads labeled

with anti-digoxigenin to produce "dumbbells," as shown in Figure 1A.

Optical tweezers experiments

The single-molecule manipulation of DNA hairpin was done on a dual-beam mini-tweezers dual-trap optical tweezers, described in detail previously [24]. The DNA was held between two polystyrene beads (Spehrotech, Libertyville, IL) by immobilizing the labeled DNA onto the surface of beads. The 2-μm diameter polystyrene bead was positioned by suction on the tip of a micropipette, which was fixed in the chamber and coupled to piezoelectric flexure stage for small displacements. When a single DNA molecule was tethered between the two beads, force-extension measurements were made by measuring the force on the bead in the trap while moving the trap with a loading speed of 200 nm/sec. The position measurements were converted to a measurement of the molecular extension by correcting for the trap stiffness. Experiments were performed in a buffer solution containing 20mM HEPES, pH 7.5, 150mM KCl, 2 mM $MgCl_2$. To analyze the unfolding behavior of DNA hairpin quantitatively, the data points between 1 to 30 pN of the first stretching trace were used in the fit to get DNA elasticity parameters. For the dsDNA handle, we used worm-like chain model (WLC) to do the fitting:

$$\frac{x}{L} = 1 - \frac{1}{2}\left(\frac{k_B T}{FP}\right)^{\frac{1}{2}} + \frac{F}{K}$$

where x is the end-to-end distance, k_B is the Boltzmann constant, T is the temperature at 298 K, P is the persistence length, F is the force, and K is the elastic stretch modulus. Here we used P= 50 nm, L=0.34 nm/bp, K=1200 pN. For the contour length change of the single-stranded DNA, we used FJC model to do the fitting:

$$\frac{x}{L} = [\cos(\frac{2FP}{k_B T}) - \frac{1}{2}\left(\frac{k_B T}{FP}\right)](1 + \frac{F}{K})$$

Unfolding and refolding traces were individually fitted using P= 1 nm, L=0.55 nm/nt, K=800 pN.

Single-molecule FRET experiments

All DNA oligonucleotides used for single-molecule experiments were purchased from Invitrogen (Thermo Fisher Scientific, Shanghai). DNA targets were prepared by mixing the two complement strands and heating to 95 °C for 5 min followed by cooling to room temperature over 1 h. Cy3/Cy5-labelled DNA targets were immobilized on the polyethylene glycol-passivated cover-glass surface using streptavidin–biotin interaction. We acquired fluorescence data using an objective-based TIR microscope. Imaging was performed at room

temperature in a buffer (20 mM HEPES, pH 7.5, 100 mM KCl, 2 mM $MgCl_2$. An oxygen scavenging system (2 units μl^{-1} glucose oxidase, 20units μl^{-1} catalase, 0.8% β-D-glucose and 2 mM Trolox (Sigma-Aldrich)) was used in all experiments to prevent the organic fluorophores from severe photo-fatigue. The time resolution for all the experiments was 100 ms. Detailed methods of smFRET data acquisition and analysis were described in previous study [35]. The FRET efficiency of a single molecule was approximated as $FRET=I_A/(I_D+I_A)$, where I_D and I_A are the background and leakage-corrected emission intensities of the donor and acceptor, respectively. The histograms of the FRET upon Cas9-sgRNA binding were obtained by averaging the first 50 frames of each FRET trace for every individual molecule after manually filtering photo-bleaching effects.

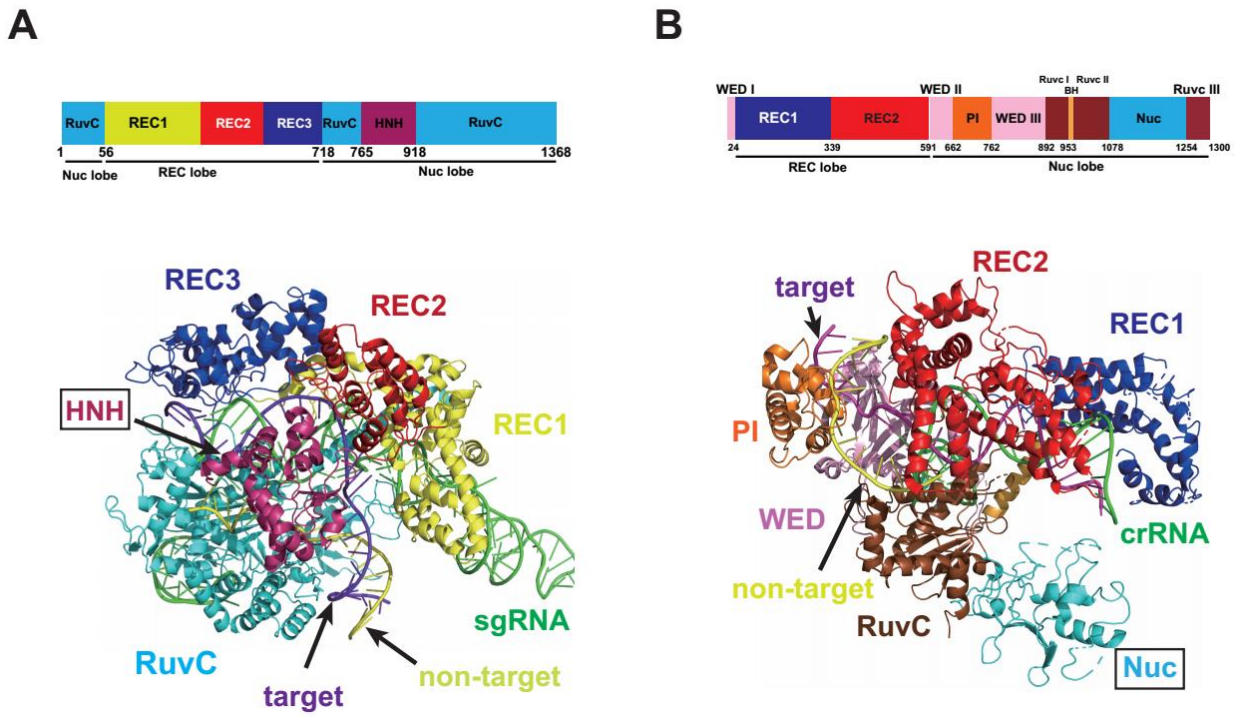

Figure S1. (A) The domain architecture (top) of SpyCas9 and its ternary complex structure with sgRNA and target DNA (bottom, PDB 5F9R); (B) The domain architecture (top) of FnCas12a and its complex structure with crRNA and target DNA (bottom, PDB 5MGA). Both Cas9 and Cas12a adopt bilobed architecture. Specifically, Cas12a consists REC and NUC lobes which are connected by the wedge (WED) domain, the N-terminal REC lobe contains two α-helical domains (REC1 and REC2) that participate in nucleic acids recognition, the crRNA-target DNA hybrid duplex bound to the central channel between the two lobes. The C-terminal NUC lobe consists of the PAM interacting (PI) domain, the RuvC and Nuc domains involved in the recognition and cleavage of the target DNA.

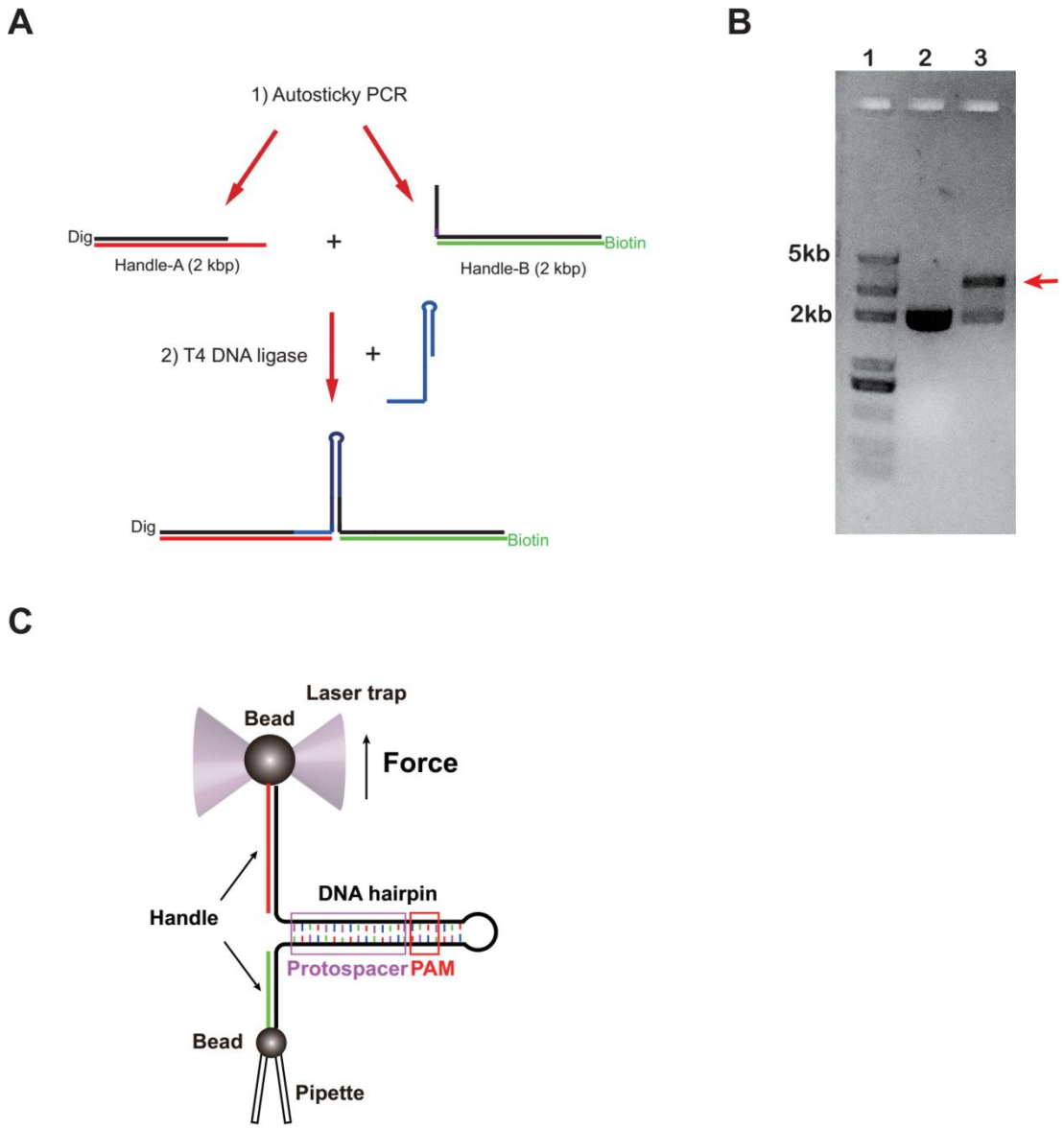

Figure S2. (A) Schematic representation of hairpin DNA construction for single-molecule experiments. (B) Gel shift assay showing the success of making hairpin construct. Lane 1, DNA ladder; lane 2, DNA handle mix; lane 3, ligated product, the ligated hairpin construct with handles on both side is indicated with red arrow. (C) The optical tweezers setup with the ligated product involving the designed DNA hairpin was tethered between the two beads, the DNA hairpin consists a PAM sequence followed by the target sequence of dCas12a-crRNA complex. One bead is sucked and immobilized by a micropipette while the other bead is pulled by a laser trap.

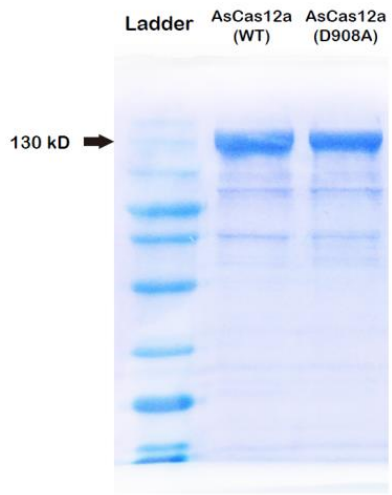

Figure S3. SDS-polyacrylamide gel electrophoresis of AsCas12a and its dead mutant D908A (dCas12a) used for the single-molecule experiments. The dead mutant has similar quality as the WT one.

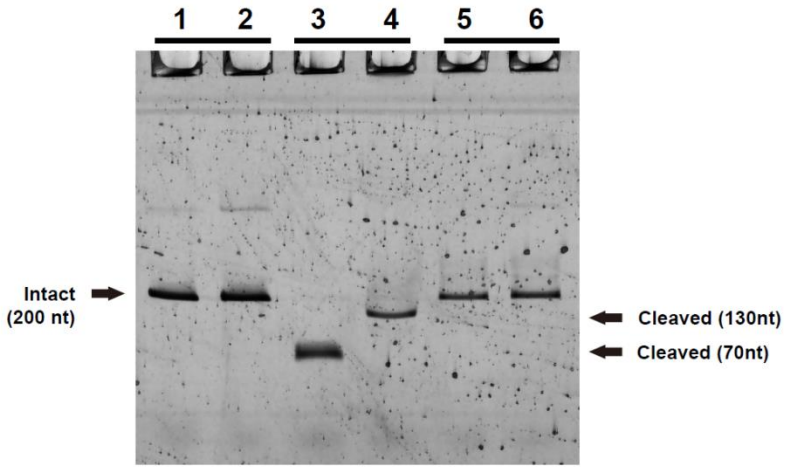

Figure S4. Electrophoretic mobility shift analysis of the target DNA cleavage using fluorophore-labeled target DNA duplexes. Lane 1, 3, 5, target strand labeled with Alexa-488; Lane 2, 4, 6, non-target strand labeled with Alexa-488; Lane 1,2: DNA alone; Lane 3,4: DNA incubated with Cas12a-crRNA complex; Lane 5, 6: target DNA incubated with Cas12a(D908A)-crRNA complex. The result confirmed that the dead mutation (D908A) eliminates the cleavage activity of Cas12a-crRNA complex.

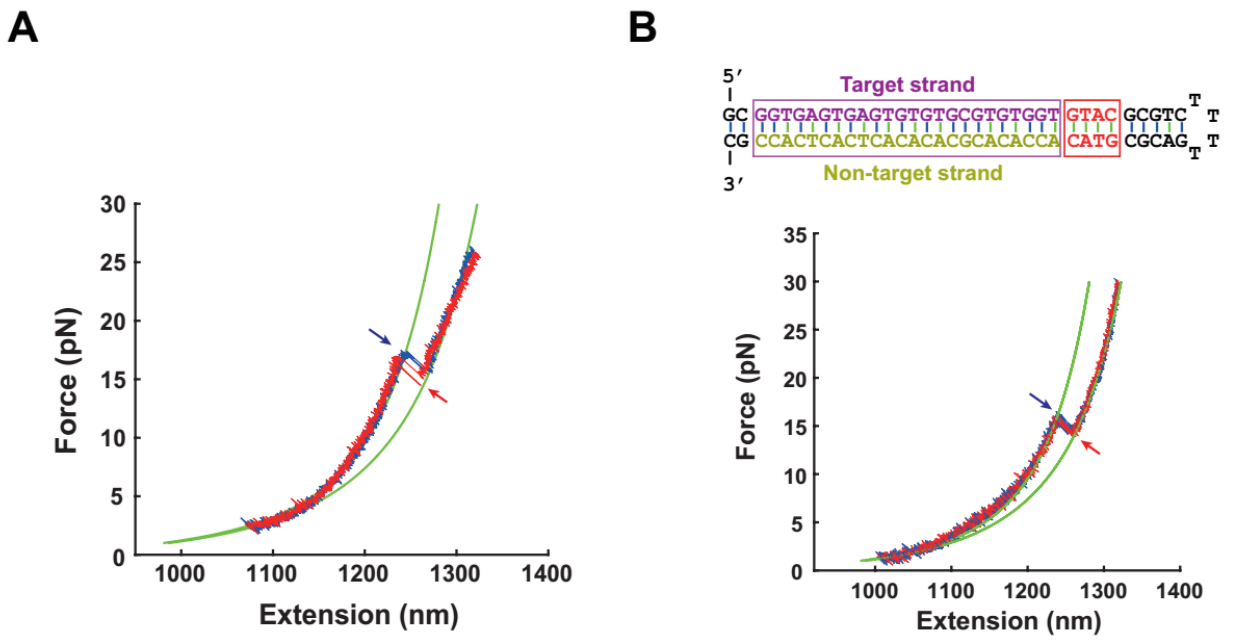

Figure S5. (A) Stretching (blue) and relaxation (red) force-extension curves from DNA hairpin containing the target DNA sequence in the presence of crRNA only, but without Cas12a. (B) The sequence of DNA hairpin containing target sequence with a mutated PAM sequence (top), the representative stretching (blue) and relaxation (red) force-extension curves with this PAM-mutated DNA hairpin the presence of Cas12a-crRNA complex was shown (bottom). Only one transition can be observed for PAM-mutated sequence, suggesting that R-loop formation requires proper Cas12a-crRNA and target DNA recognition.

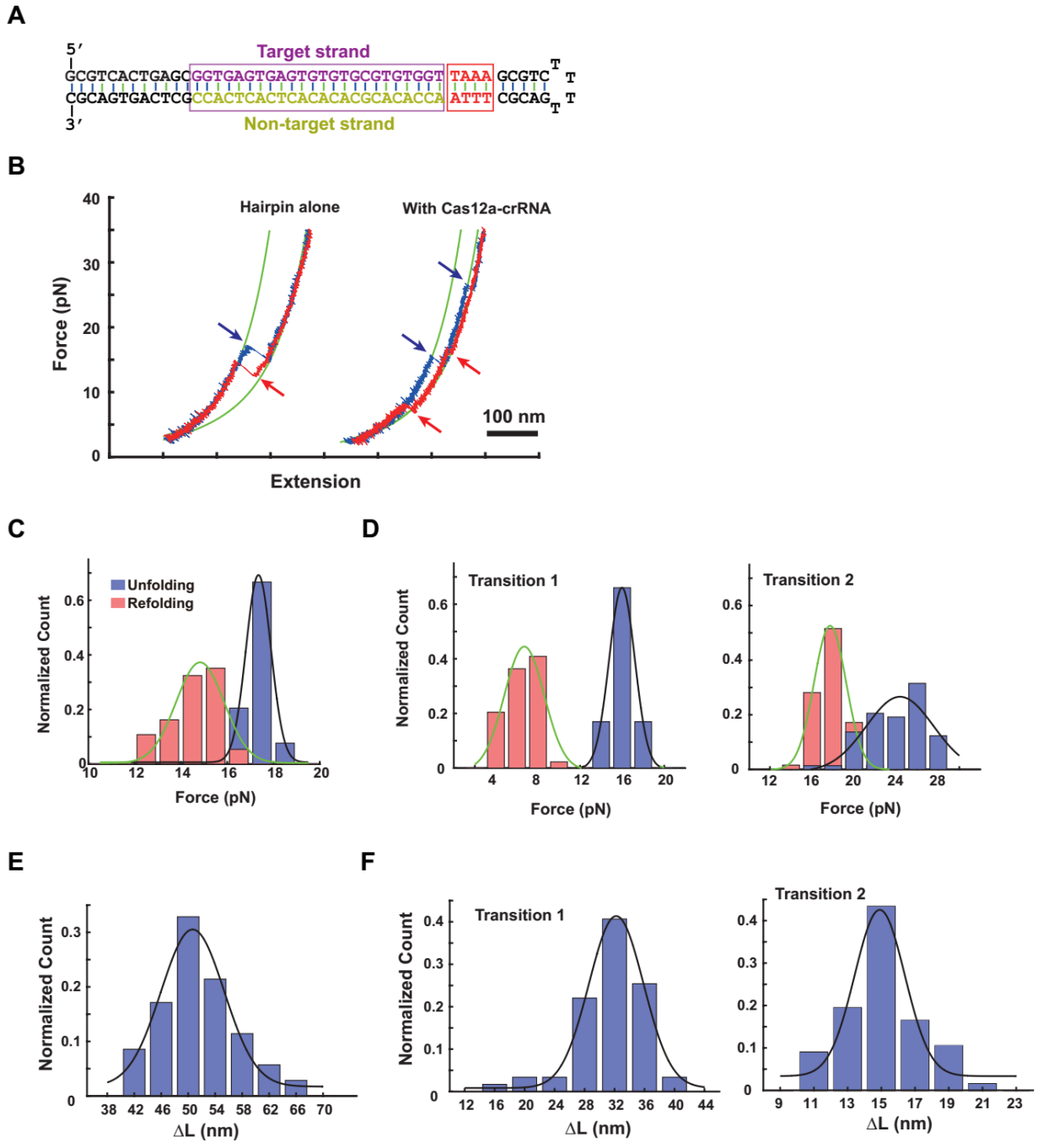

Figure S6. (A) The sequence of target DNA hairpin with an extra 10-bp outside the PAM distal region. (B)The representative stretching (blue) and relaxation (red) force-extension curves for this DNA sequence in the absence (left) and presence of Cas12a-crRNA complex (right) are shown. (C) Histogram of the Unfolding (blue) and refolding (red) forces for DNA hairpin alone, n=228. The relative larger difference between the unzipping force (17.4 ± 0.5 pN) and zipping force (14.8 ± 0.5 pN) for this longer DNA hairpin than that of shorter one (Figure

1) indicates that the folding rates of the DNA hairpin with longer stem length is slower than that of shorter stem length.[27] (D) Histogram of the Unfolding (blue) and refolding (red) forces for R-loop complex during the first (T1, left) and second (T2, right) transitions, n=107. (E) Histogram of the contour length change (ΔL) for the DNA hairpin alone. (F) Histogram of the contour length change (ΔL) for R-loop complex during the first (Transition 1 or T1, left) and second (Transition 2 or T2, right) transitions.

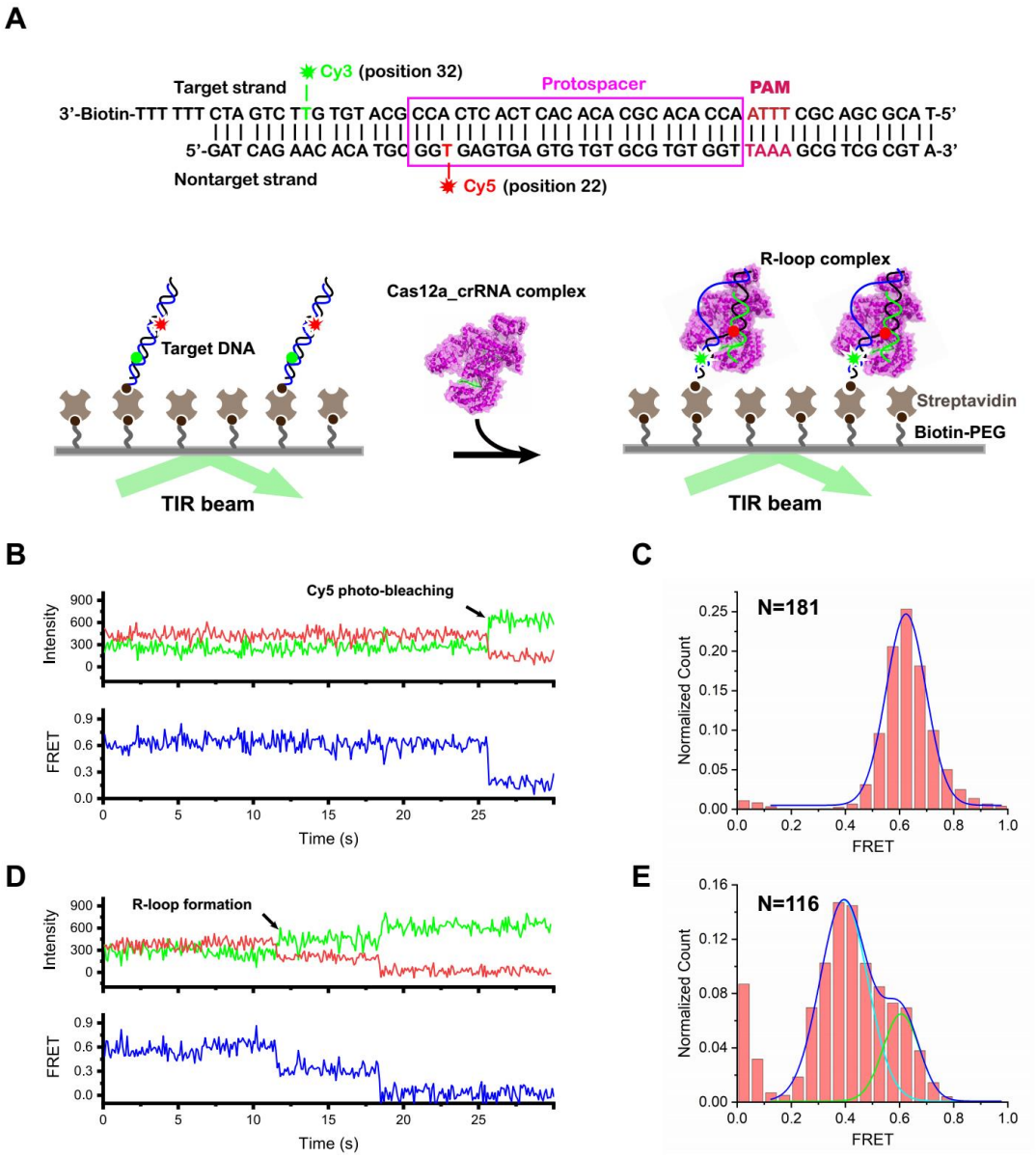

Figure S7. smFRET study on the R-loop formation. (A) Schematic representative of smFRET experimental setup. Biotinylated target DNA was labeled with Cy3 (donor) at +32 position on the target strand and Cy5 (acceptor) fluorophores at +22 position on the nontarget strand, respectively. The target DNA was first immobilized on the biotin functionalized glass surface of microfluidic chamber through biotin-streptavidin interaction. Progression of DNA binding and R-loop formation was monitored by measuring the FRET signal changes between Cy3 and Cy5. (B)-(C) Typical smFRET trace (B) and FRET histogram (C) for target DNA

without Cas12a–CrRNA complex. (D)-(E) Typical smFRET trace (D) and FRET histogram (E) for target DNA with Cas12a–CrRNA complex.

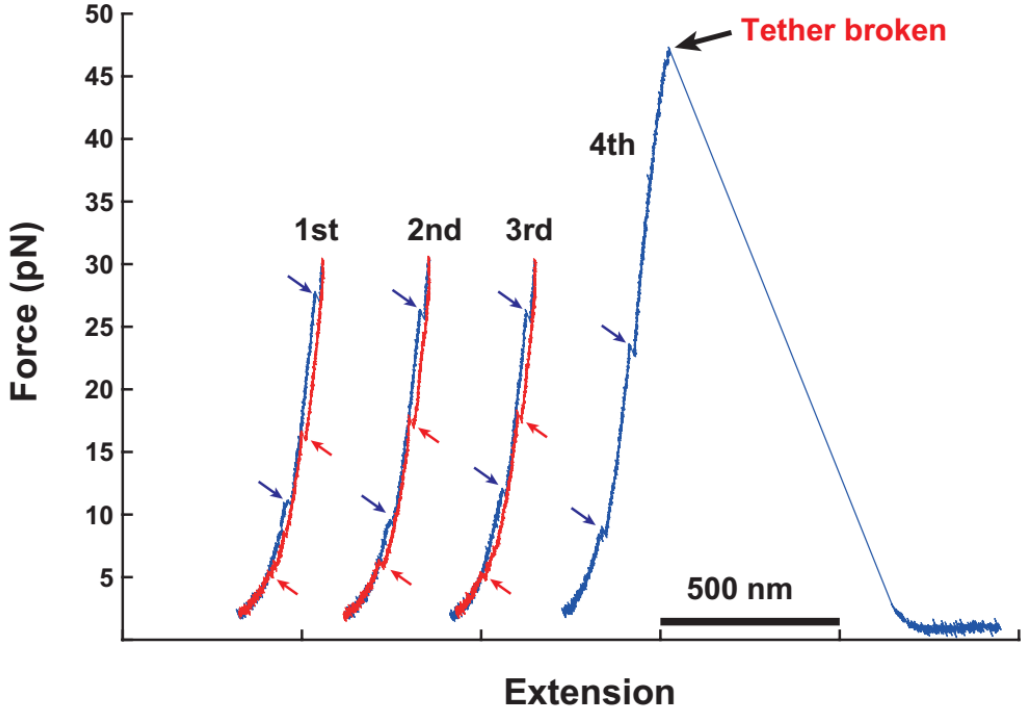

Figure S8. Consecutive force loading and relaxation of the DNA hairpin in the presence of Cas12a-crRNA complex. Two-transition pattern can be consecutively observed for the first three experimental cycles. During the fourth pulling, the force is constantly increased until the tether is broken. Blue arrows indicate unzipping transitions and red arrows indicates zipping transitions. These results indicated that Cas12a-crRNA complex will not leave the system, correct Cas12a-crRNA-target DNA complex can be formed with RNA/DNA hybrid and R-loop.

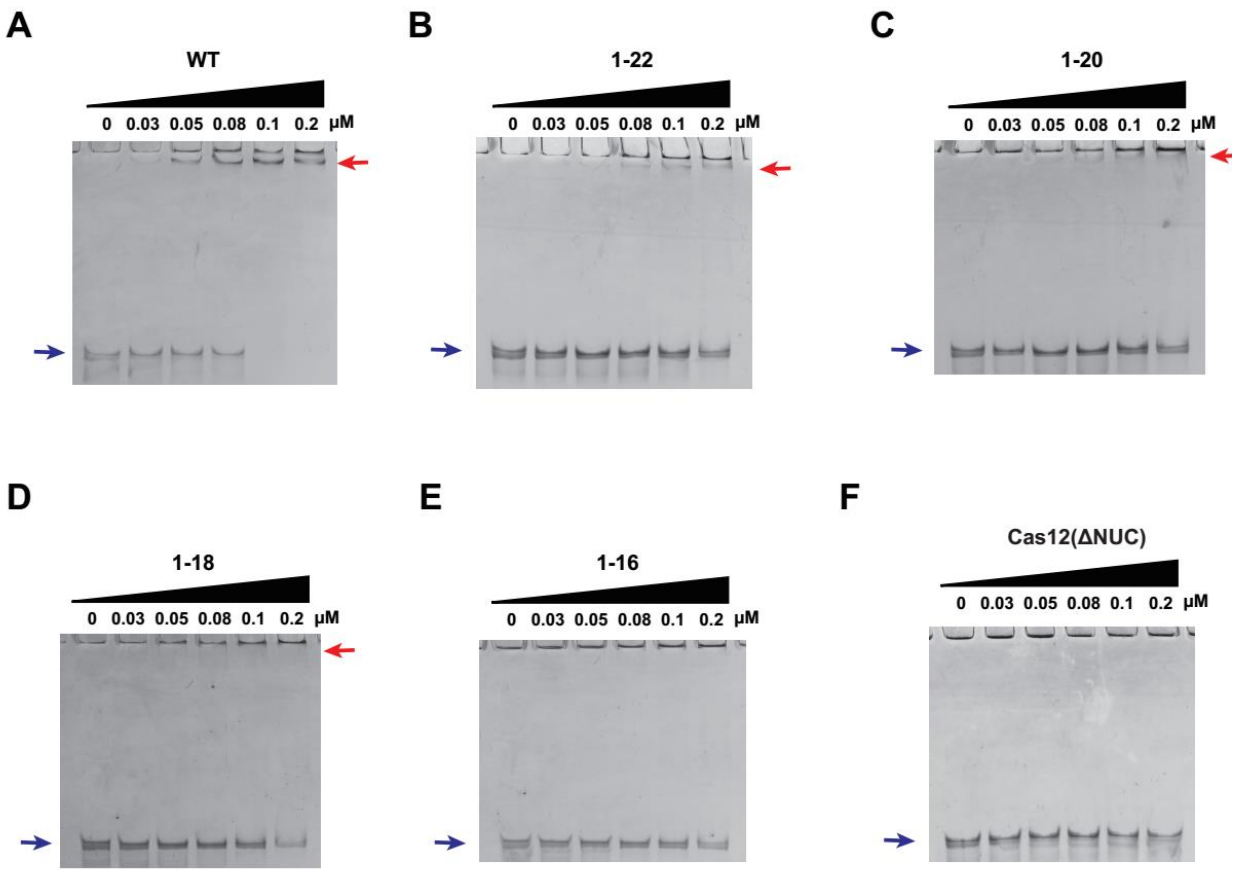

Figure S9. (A)-(E) Electrophoretic mobility shift assay to study the target binding by using catalytically inactive dCas12a–crRNA complexes and fluorophore-labeled target DNA duplexes. The sequences of the guide region of the crRNA are the same as those of Figure 2C. (F) Target binding assay for WT sequence of crRNA using dCas12a(ΔNuc) mutant. Blue arrows indicate DNA alone and red arrows indicate R-loop complex.

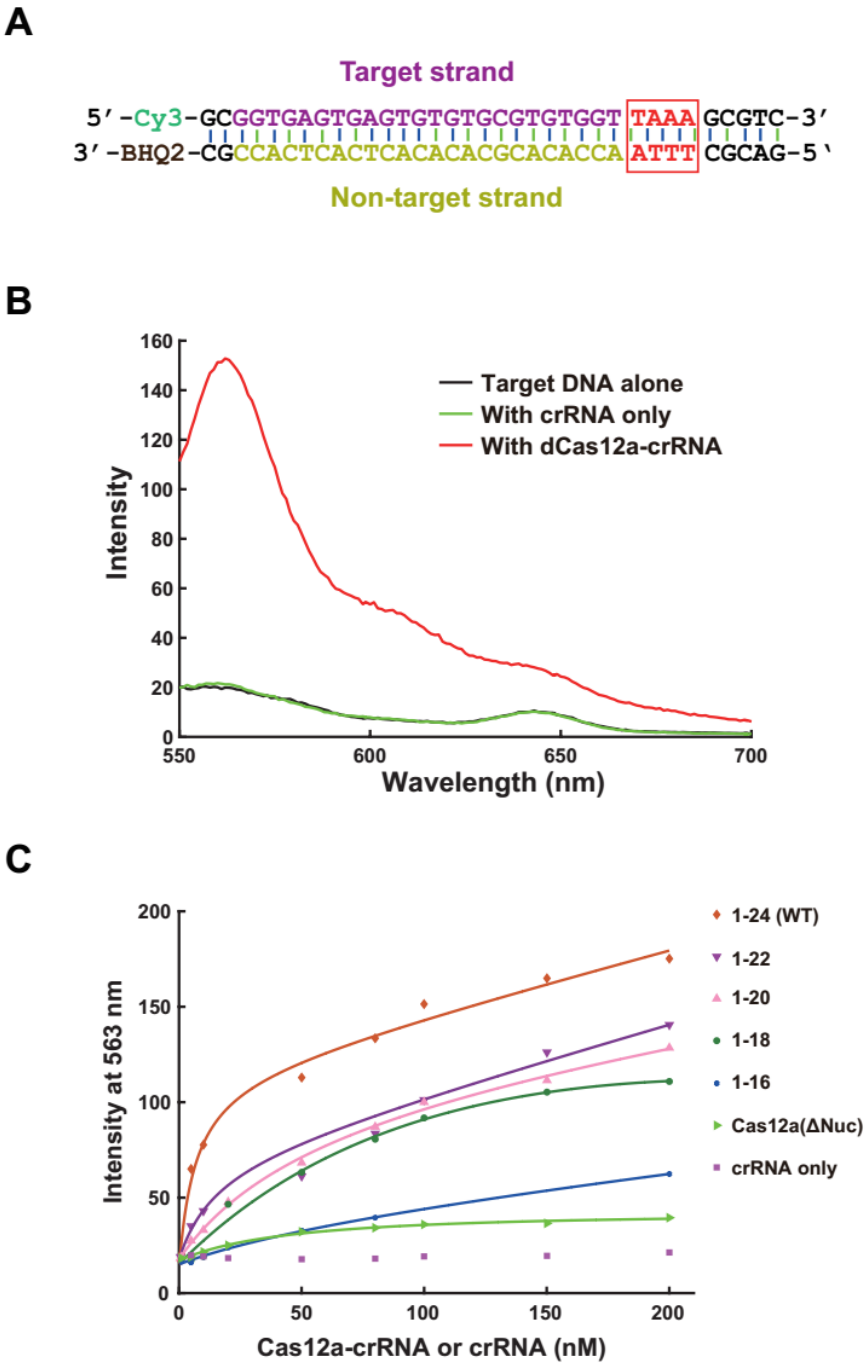

Figure S10. A FRET-based target DNA binding assay with dCas12a-crRNA complex. (A) The sequence of dual labeled target DNA duplex used as the substrate for the Cas12a binding. The 5' end of the target strand was labeled with Cy3 and the 3' of the non-target strand was labeled with BHQ2. BHQ2 can quench the emission of Cy3 when they are close to each other. (B) The fluorescence emission spectra of the DNA duplex (50 nM) in the absence or presence of the crRNA or Cas12a-crRNA complex (100 nM). The Cy3 fluorescence intensity stays

at low level for the target DNA alone or in the presence of crRNA because the DNA duplex is well formed and the Cy3 fluorescence is quenched by BHQ2. In the presence of dCas12a-crRNA, BHQ2 was away from Cy3 after the R-loop formation and duplex opening, fluorescence intensity increases at much higher level. (C) Target DNA binding assay with dCas12a harboring crRNAs containing 0, 2, 4, 6, 8-bp mismatches at the PAM-distal end and dCas12a(ΔNuc) harboring WT crRNA. Fittings to the experimental data (shown as solid lines) yield equilibrium dissociation constants (Kd) of 6.5, 14.6, 17.8, 47.9 nM, and 76.8 nM, respectively. The Nuc truncated dCas12a (dCas12a(ΔNuc)) shows significant lower binding affinity with its target (Kd =76.8 nM) than that of wild-type dCas12a.

Table S1. The unfolding force (F_{un}) and contour length change (ΔL) of the two hairpins in the absence and presence of dCas12a-crRNA complex.

Name	Without dCas12a-crRNA		With dCas12a-crRNA			
	F_{un} (pN)	ΔL (nm)	Transition 1		Transition 2	
			F_{un} (pN)	ΔL (nm)	F_{un} (pN)	ΔL (nm)
74-nt hairpin	17.8 ± 0.5	38.5 ± 0.5	9.5 ± 0.3	15.3 ± 0.3	25.3 ± 0.2	15.1 ± 0.2
94-nt hairpin	17.7 ± 0.2	50.7 ± 0.5	16.0 ± 0.2	32.5 ± 0.2	24.4 ± 0.5	14.9 ± 0.2

Table S2. Sequences of crRNA used in this study. Red indicates that base-pairing region, blue indicates mismatches.

Name	sequence, 5'- 3'
crRNA-WT	GGUAAUUUCUACUCUUGUAGAUACCACACGCACACACUCACUCACC
crRNA-$M_{1\text{-}22}$	GGUAAUUUCUACUCUUGUAGAUACCACACGCACACACUCACUCAGG
crRNA-$M_{1\text{-}20}$	GGUAAUUUCUACUCUUGUAGAUACCACACGCACACACUCACUGUGG
crRNA-$M_{1\text{-}18}$	GGUAAUUUCUACUCUUGUAGAUACCACACGCACACACUCAGAGUGG
crRNA-$M_{1\text{-}16}$	GGUAAUUUCUACUCUUGUAGAUACCACACGCACACACUGUGAGUGG