ISCI, Volume *3*

Supplemental Information

Structural and Mechanistic Analyses Reveal a Unique Cas4-like Protein in the Mimivirus Virophage Resistance Element System

Chao Dou, Mingjing Yu, Yijun Gu, Jinjing Wang, Kun Yin, Chunlai Nie, Xiaofeng Zhu, Shiqian Qi, Yuquan Wei, and Wei Cheng

Supplemental Information

Supplemental Figures

Figure S1

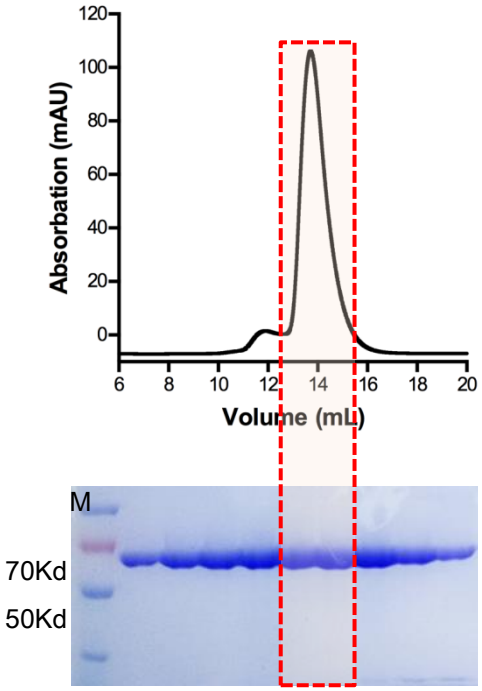

Figure S1. High quality protein purification of R354, related to Figure 2.

The upper panel indicates protein purification by gel filtration. The lower panel shows SDS-PAGE detection by Coomassie Brilliant Blue staining

Figure S2

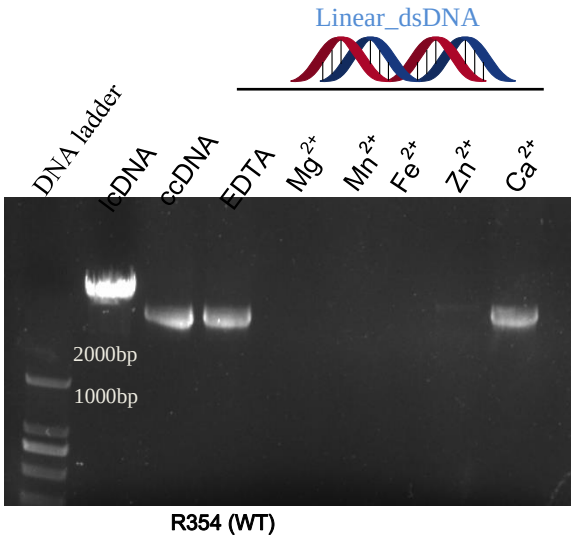

Figure S2. Divalent metal-dependent nuclease activity of R354, related to Figure 2. Characterization of the metal-dependent nuclease activity by mixing with different divalent metal ions and addition of EDTA as a negative control while performing endonuclease activity assay *in vitro.*

Figure S3

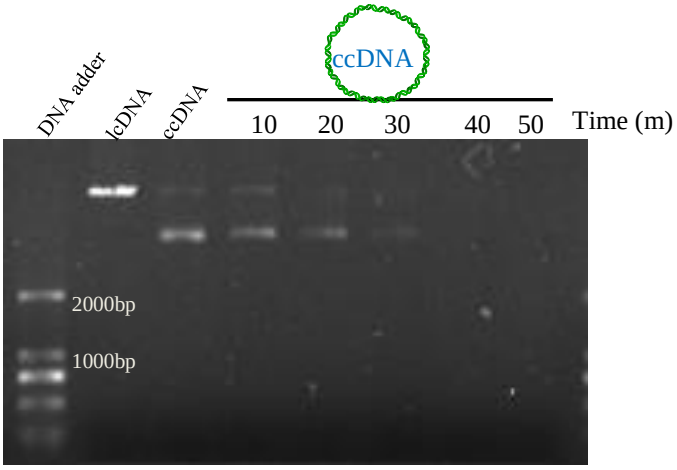

Figure S3. Nuclease activity of R354 (Δ1-128) degrading ccDNA, related to Figure 2. Nuclease activity assays showed that the activity of the NTD-deletion of R354 (Δ1-128) as well as the wild-type protein (Figure 2A), which indicated NTD is not required for the enzyme activity. CcDNA refers to circular coiled-coil DNA.

Figure S4

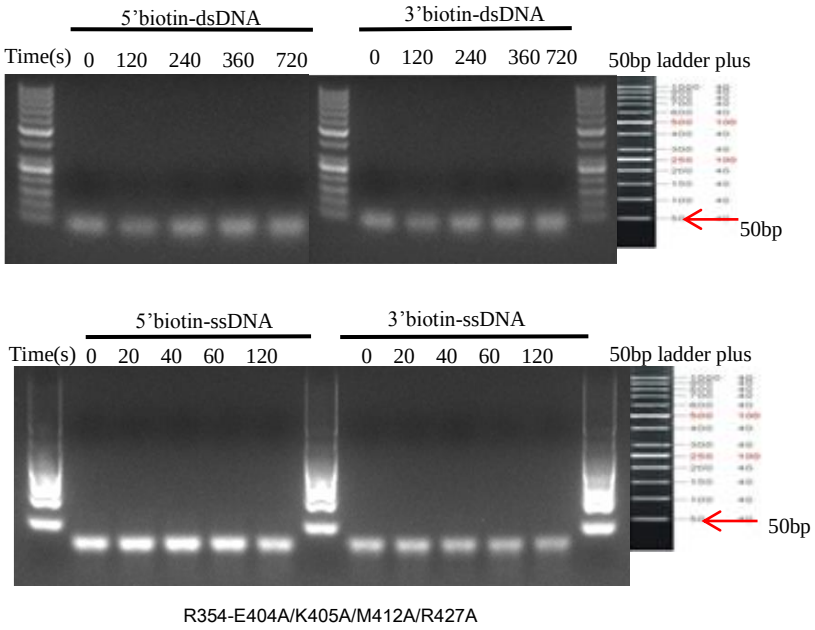

Figure S4. Exo-nuclease activity of R354-E404A/K405A/M412A/R427A, related to Figure 4.

Exonuclease activity assay using biotin-labeled dsDNA (upper panel) and ssDNA (lower panel). R354 was incubated with the labeled substrates with blunt ends or 5'- or 3'-overhangs, and the reactions were analyzed by agarose gel electrophoresis (AGE). Results indicate no detectable cleavage of the dsDNA and ssDNA.

Figure S5

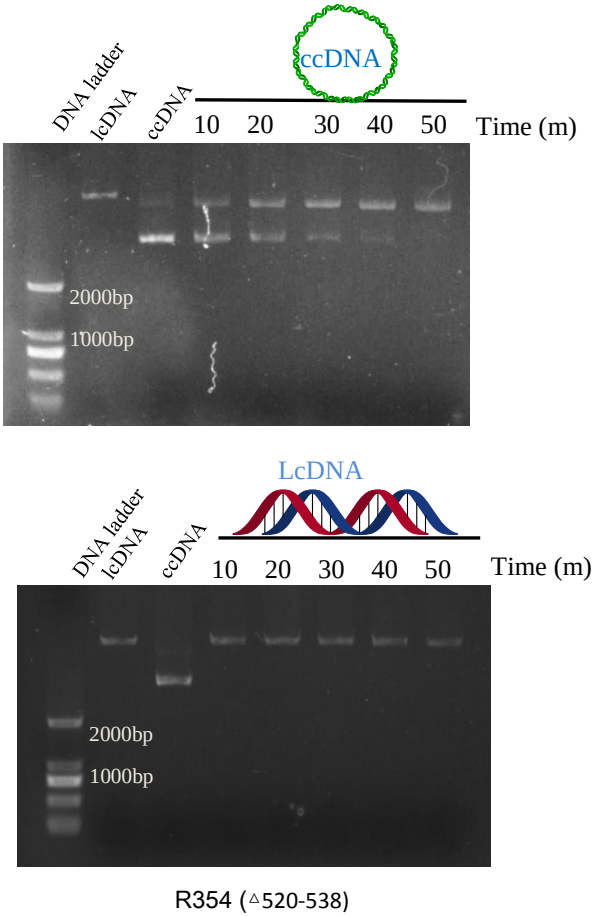

Figure S5. Nuclease activity assay for the R354 C-terminal-deletion variant, related to Figure 5. CcDNA refers to circular coiled-coil DNA; LcDNA refers to linear coiled-coil DNA.

Figure S6

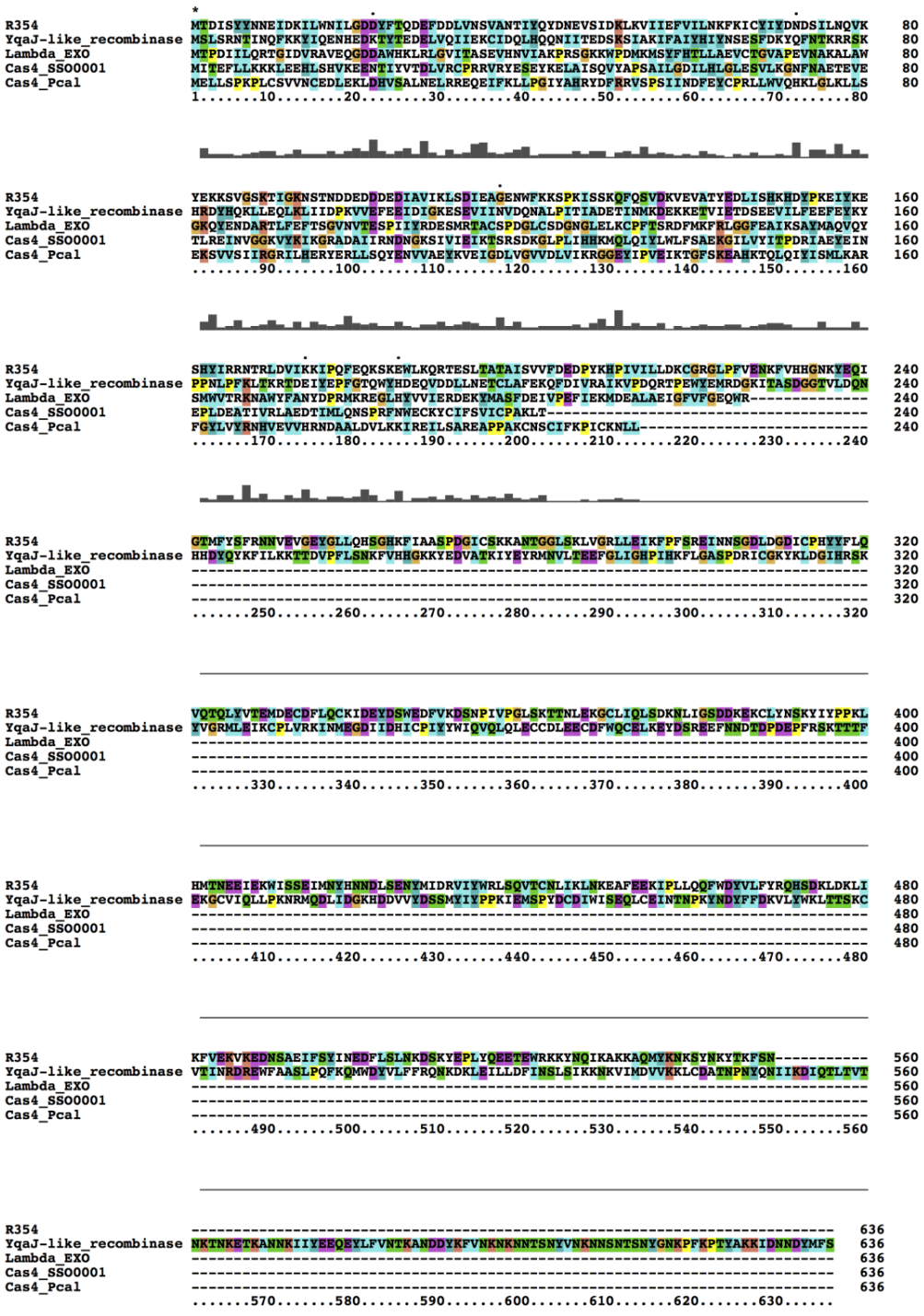

Figure S6. Amino acid sequence alignment of R354, Cas4, and λ exonulease, related to Figure 6.

Sequence alignment of R354 (*Acanthamoeba polyphaga mimivirus* [APMV]), YqaJ-like recombinase (*Klosneuvirus*), Lambda Exo (*Escherichia virus lambda*), Cas4_SSO0001 (*Sulfolobus solfataricus*), and Cas4_Pcal (*Pyrobaculum calidifontis*).

Figure S7

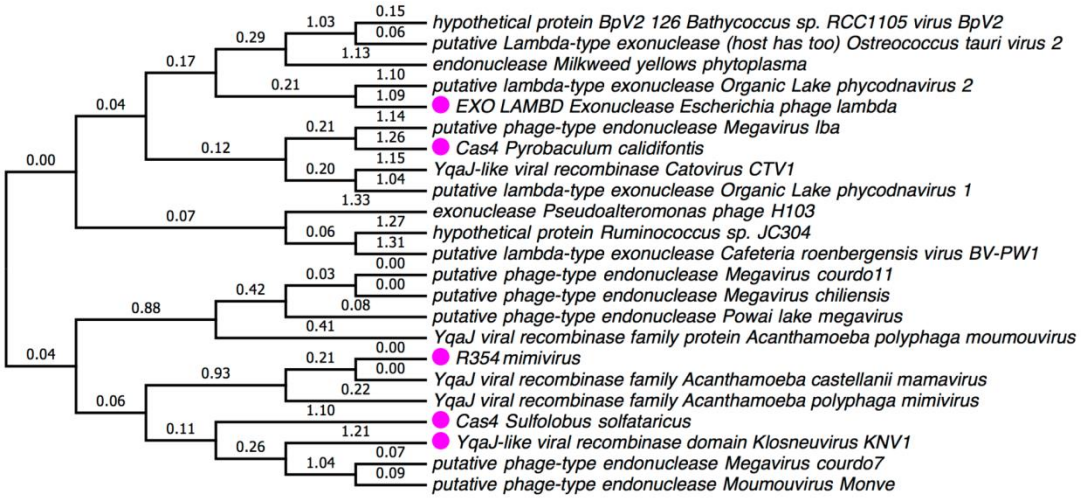

Figure S7. Phylogenetic tree distribution of representative members of the nuclease family, related to Figure 6.

A maximum likelihood phylogenetic tree of R354 was constructed using MEGA 6 software with the aligned protein sequences obtained from NCBI BLAST. Homologous proteins are labeled in magenta.

Table S1. Data collection and refinement statistics, related to Figure 3

	Native R354 (Δ1-128) PDB ID: 5YET	SeMet-R354 (Δ1-128)	Four-point mutations R354 (Δ1-128) PDB ID: 5YEU
Data collect			
Space group	P21	P21	P212121
Cell dimensions			
a, b, c (Å)	97.85, 57.81, 99.95	97.26, 57.75, 98.46	79.25, 93.45,123.19
α, β, γ (°)	90.00, 108.39, 90.00	90.00, 108.71, 90.00	90.00, 90.00, 90.00
Resolution (Å)	50-2.80 (2.85-2.8)	50-3.20 (3.26-3.20)	50-3.00 (3.05-3.00)
R_{merge} (%)	14.6	18.8	24
R_{pim} (%)	6.7	7.0	6.8
I/σ (*I*)	13.05 (2.07)	20.7(2.75)	10.33 (1.56)
Completeness (%)	98.9 (93.9)	99.8 (98.1)	100 (100)
Redundancy	4.8 (4.0)	7.2 (6.7)	12.6 (10.1)
Refinement			
No. reflections	22530		18841
R_{work}/R_{free} (%)	19.45/24.16		23.05/26.71
No. atoms	6667		6374
Protein	6617		6347
Ligand/ion			4
Average B-factors	39.27		59.32
R.m.s deviations			
Bond length (Å)	0.002		0.002
Bond angles (°)	0.495		0.579

Value in parentheses is for the highest-resolution shell.

SUPPLEMENTAL Materials

Recombinant DNA

pGEX6p-1-GST-R354 (Full-length)

pET28b-R354 (Full-length)

pET28b-R354 (△1-128)

pET28b-R354 (△520-538)

pET28b-R354-Q177A

pET28b-R354-W182A

pET28b-R354-R186A

pET28b-R354-K205A

pET28b-R354-K225A

pET28b-R354-V227W

pET28b-R354-D268A

pET28b-R354-E289A

pET28b-R354-K291A

pET28b-R354-R296A

pET28b-R354-Y312A

pET28b-R354-Y313A

pET28b-R354-Q316A

pET28b-R354-Q177A/K389A

pET28b-R354-Q177A/K389A/Y392A/D418A

pET28b-R354-E234A/Q235A/E251A

pET28b-R354-D268A/E289A/K291A

pET28b-R354-E404A/K405A/M412A/R427A

Sequence-based reagents

dsDNAs biotin labeled at 5' end:

Biotin-5'TTTTGGCTCGTGGAGACCATCTTTCTTTCAAGAGAAAGATGGTCTCCACGAG

CC-3'

3'-CCGAGCACCTCTGGTAGAAAGAAAGTTCTCTTTCTACCAGAGGTGCTCGGTTTT

5'-Biotin

dsDNAs biotin labeled at 3' end:

5'-GGCTCGTGGAGACCATCTTTCTTTCAAGAGAAAGATGGTCTCCACGAGCCTTTT-3'

-Biotin

Biotin-3'TTTTCCGAGCACCTCTGGTAGAAAGAAAGTTCTCTTTCTACCAGAGGTGCTC

GG-5'

ssDNAs biotin labeled at 3' end:

Biotin-5'-TATTAGATAAATGGGCAAGTTTGTGGAAATCGGCAAGGTGTTATATAAATAT

AAAGTAG-3'

ssDNAs biotin labeled at 3' end:

5'-TATTAGATAAATGGGCAAGTTTGTGGAAATCGGCAAGGTGTTATATAAATATAAAG

TAG-3'-Biotin.

Transparent Methods

Cloning, protein expression, and purification

A codon-optimized gene encoding full-length R354 and an N-terminal 129 amino acid deletion fragment from mimivirus were subcloning into the vectors pET-28b (Novagen, Madison, WI, USA), and pGEX-6P-1 (GE Healthcare), respectively.

All expression plasmids were transformed into *E. coli* BL21 (DE3) after induction with 300 μM IPTG at 22°C for 18 h. Cell pellets harvested by centrifugation were resuspended in buffer containing 300 mM NaCl and 25 mM Tris-HCl pH 8.0. After sonication, the cell lysate containing 6x His-tagged protein was applied to a nickel-affinity column (Ni-NTA; GE Healthcare). The column was washed using suspension buffer plus 20, 30, and 40 mM imidazole successively. Subsequently, purification was assessed using anion exchange chromatography (Source Q; GE Healthcare) and size-exclusion chromatography (Superdex200 10/300; GE Healthcare). After gel filtration, the peak fractions were assessed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and flash-frozen in liquid nitrogen for crystallization. GST fusion proteins were purified using glutathione-coated sepharose resin (GE Healthcare) for glutathione S-transferase (GST) pull-down assays. Selenomethionine (SeMet)-substituted R354 was expressed and purified similarly to native R354.

Protein crystallization

Native R354 was concentrated to 8 mg/mL, and the mature native crystal was

screened in a buffer containing 200 mM $MgCl_2$ (pH 5.9) and 20% PEG3350. Other crystals of 20 mg/mL NTD (129 amino acid-long) and 8 mg/mL SeMet were obtained in a solution of 200 mM ammonium tartrate (pH 6.6-7.0) and 20% PEG3350. All R354 proteins were crystallized at 20°C. Crystals grew to full size in 10 days and were flash-frozen in liquid nitrogen with additional 10% glycerol as a cryo-protectant before X-ray diffraction.

Data collection and structure determination

All datasets were obtained at beamlines BL17U1and BL19U1 at the Shanghai Synchrotron Radiation Facility (SSRF). The collected frames were processed and scaled using the XDS data processing package (WANG et al., 2015) and HKL2000 (Zhang et al., 2011). The native crystals diffracted to 3.0 Å in the P21212 space group and the 4-point mutation crystals diffracted to 3.1 Å in the P21212 space group. The crystals of selenomethionine (SeMet)-substituted R354 protein diffracted to 2.8 Å, and the phases were determined by the single-wavelength anomalous dispersion (SAD) method. The final model was manually built in COOT (Emsley et al., 2010) and refined in Phenix2 (Afonine et al., 2012).

DNA annealing and ITC

The substrates consisting of 5 oligonucleotides (AT-rich substrates) were synthesized by the Tsingke Company (Chengdu , China) and purified by high-performance liquid chromatography (HPLC). The dsDNAs were obtained using

an annealing program as follows: equal amounts of sense and antisense ssDNA were mixed together and incubated at 98°C for 10 min, followed by 60 min at 15°C.

The annealed DNA prepared for the ITC assay was dissolved in the same buffer used for storing R354, which contained 10 mM HEPES pH 8.0 and 100 mM NaCl. The mutated R354 proteins were purified according to a routine protocol and stored in buffer containing 10 mM HEPES pH 8.0 and 100 mM NaCl. The concentrations of the mutated R354 proteins and dsDNA were 0.03 and 0.45 mM, respectively. ITC experiments were performed on a MicroCal iTC200 calorimeter (GE Healthcare) using the following settings: total injections, 19; cell temperature, 25°C; reference power, 5 μcal/s; initial delay, 60 s; stirring speed, 750 rpm; injection volume, 2 μL; and spacing time, 150 s. The original titration datasets were processed using Origin 7.0 software.

In vitro gel-based enzyme activity

Endonuclease enzymatic activity

The plasmid pET21b-LipL (6660 bp) was constructed by sub-cloned the gene LipL (GeneID: 886575) into the vectors pET-21b (Novagen, Madison, WI, USA), to linearize the circular coiled-coil DNA, the plasmids were digested by XhoI.

Gel-based 3'-5' dsDNA enzyme activity

The 5**'** and 3**'** biotin-labeled dsDNAs and ssDNAs were synthesized by the Xiamen Niuketai Biological Technology Co. Ltd., (Xiamen, China).

dsDNAs biotin labeled at 5' end:

Biotin-5'TTTTGGCTCGTGGAGACCATCTTTCTTTCAAGAGAAAGATGGTCTCCAC

GAGCC-3';

3'-CCGAGCACCTCTGGTAGAAAGAAAGTTCTCTTTCTACCAGAGGTGCTCGGTT

TT 5'-Biotin;

dsDNAs biotin labeled at 3' end:

5'-GGCTCGTGGAGACCATCTTTCTTTCAAGAGAAAGATGGTCTCCACGAGCCTTT

T-3'-Biotin;

Biotin-3'TTTTCCGAGCACCTCTGGTAGAAAGAAAGTTCTCTTTCTACCAGAGGTG

CTCGG-5';

ssDNAs biotin labeled at 5' end:

Biotin-5'-TATTAGATAAATGGGCAAGTTTGTGGAAATCGGCAAGGTGTTATATAAA

TATAAAGTAG-3';

ssDNAs biotin labeled at 3' end:

5'-TATTAGATAAATGGGCAAGTTTGTGGAAATCGGCAAGGTGTTATATAAATATAA

AGTAG-3'-Biotin;

For the dsDNA enzyme activity assays, the system (20 μL) contained 1.25 μM dsDNA substrate, 0.4 μM wild type protein, 10 mM Tris/HCl (pH 8.0), and 100 mM NaCl. All reactions were performed in PCR tubes at 37°C. Five gradients were set up to collect the products. All the reactions were terminated by the addition of 5 mM EDTA and loaded onto a 3% standard agarose gel for electrophoresis in TAE [40 mM Tris, 20 mM acetic acid, and 1 mM EDTA (pH 8.0)] buffer at 8 V/cm. Gels stained by GoldView were visualized and imaged with a Bio-Rad UV transilluminator.

Gel-based 3'-5' and 5'-3' single-strand DNA enzyme activity

The reaction system (20 μL) contained 1.25 μM labeled ssDNA substrate, 0.2 μM R354 wild type protein, 10 mM Tris/HCl (pH 8.0), and 100 mM NaCl. The assay was performed as described above. Gels were stained by with Cyber Gold (AAT Bioquest, CAT. TJ1702) and visualized and imaged with a Bio-Rad UV transilluminator.

Gel-based in vitro enzyme activity

The reaction system (20 μL) consisted of 10 ng/μL dsDNA, 0.2 mg/mL R354 protein, 10 mM Tris/HCl (pH 8.0), 100 mM NaCl, and 3 mM dithiothreitol (DTT). All components of the reaction except the enzyme were pre-incubated at 37°C for 5 min. All reactions were performed in PCR tubes, and equal volumes of phenol- chloroform were added to stop the reactions. Reactions were performed at 5 different time points (5, 10, 15, 30, 45, and 60 min). The time-gradient reaction supernatants were loaded onto a 0.8% agarose gel and electrophoresed in TAE buffer at 8 V/cm. Gels were stained with GoldView. Control lanes labeled “CC” contained an equivalent amount of DNA as the reaction lanes but did not contain the enzyme. Control lanes labeled “LC” contained the XhoI-digested linearized DNA.

Quantitative exonuclease activity.

Quantitative exonuclease assay was performed by detecting the fluorescence decrease with Molecular Devices SpectraMAX machine at 484 nm excitation, 522 nm emission. The enzyme reaction including 0.05 nM XhoI linearized lipL-pET21b plasmid (6660 bp, 0.1 nM) which was prestained with PicoGreen (1⁄10000 dilution,

Molecular Probes) in buffer containing 10 mM Tris (pH 8.0), 5 mM $MgCl_2$, and 100 mM NaCl. The reaction systems were incubated for equilibration at 37 °C for 3 min prior to being initialized by adding an excess amount of enzymes (20 nM, each of mutants and wide type). Moreover, a negative control was set without any enzyme mixing. The digestion rate in nucleotides per second was obtained from raw fluorescence decrease of the linear portion, along with the fact that at saturation both ends of the DNA duplex are digested simultaneously. Standard errors are based on values from three independent technical experiments.

GST pull-down assays

GST pull-down assays were performed to validate the interactions between the wild-type and mutant proteins. The wild type protein containing a GST tag was immobilized on glutathione-coated sepharose resin in a buffer containing 25 mM Tris-HCl (pH 8.0), 300 mM NaCl, and 3 mM DTT and incubated with purified inactive protein containing the 4 mutations at a 1:1 stoichiometry. The resin was washed with 5 column volumes of buffer and analyzed using SDS-PAGE followed by Coomassie Blue staining.

Sequence alignment

Multiple sequence alignments were generated on the ClustalW online service and were edited using the ESPript 3.0 program(Larkin et al., 2007).

Supplemental References

1. WANG, Qi-Sheng, Feng, HUANG, Sheng, ZHANG, Kun-Hao, wANG, Zhi-Jun, & Chun-yan. (2015). The macromolecular crystallography beamline of SSRF. 核技术(英文版)*, 26*(1), 8-13.

2. Zhang, J., Mccabe, K. A., & Bell, C. E. (2011). Crystal structures of λ exonuclease in complex with DNA suggest an electrostatic ratchet mechanism for processivity. Proc. Natl. Acad. Sci. U. S. A.*, 108*(29), 11872.

3. Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr.*, 66*(4), 486-501.

4. Afonine, P. V., Grossekunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., & Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr.*, 68*(4), 352–367.

5. Larkin, M., Blackshields, G., Brown, N., Chenna, R., Mcgettigan, P., Mcwilliam, H., Valentin, F., Wallace, I., Wilm, A., & Lopez, R. (2007). Clustal W and Clustal X ver. 20.