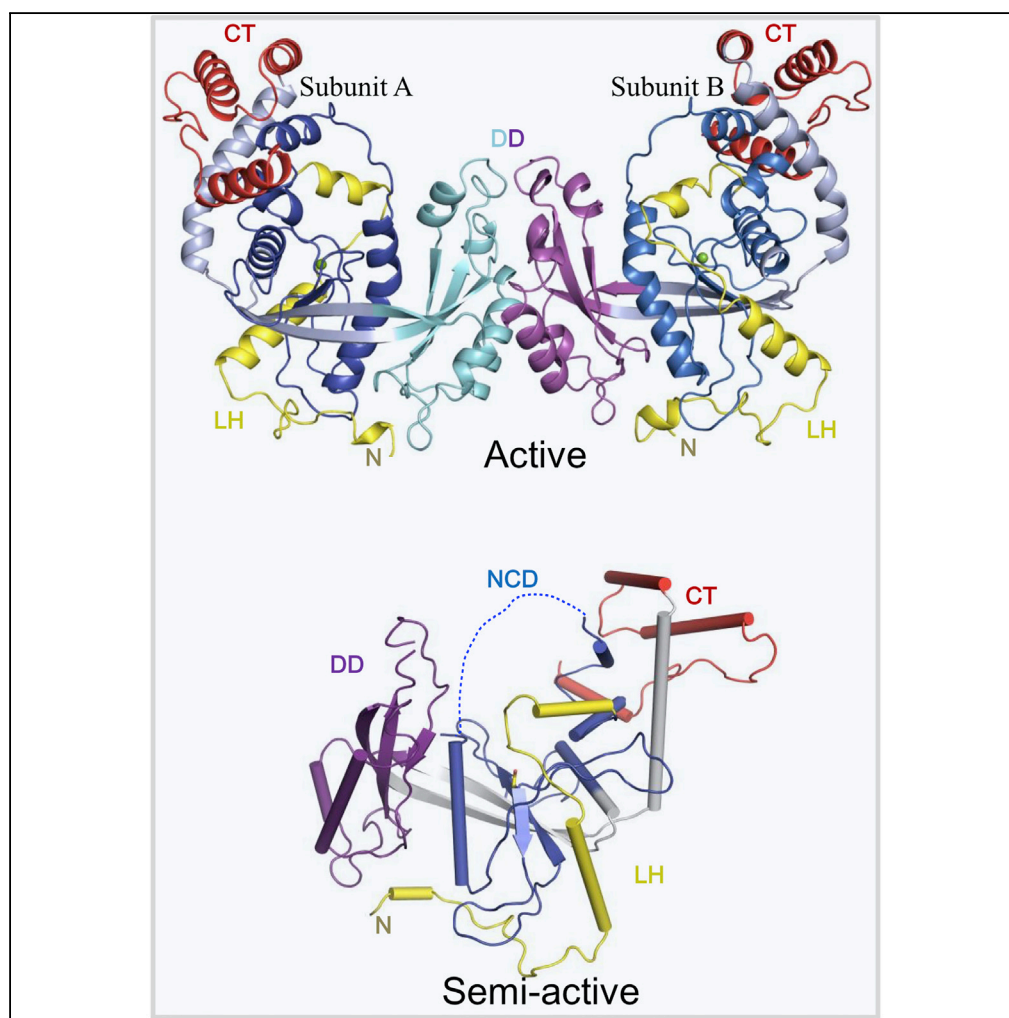**Article**

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

Chao Dou,
Mingjing Yu, Yijun
Gu, ..., Shiqian Qi,
Yuquan Wei, Wei
Cheng

chengwei669@scu.edu.cn

HIGHLIGHTS

The key component of
Cas4-like R354 in
MIMIVIRE is a dual
nuclease

Structural insights into
R354 indicate that it is fully
active as a homodimer

The semi-active state of
R354 is a monomer

DATA AND
SOFTWARE
AVAILABILITY
5YET
5YEU

Article

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

Chao Dou,[1,4] Mingjing Yu,[1,4] Yijun Gu,[2,4] Jinjing Wang,[1] Kun Yin,[3] Chunlai Nie,[1] Xiaofeng Zhu,[1] Shiqian Qi,[1] Yuquan Wei,[1] and Wei Cheng[1,5,*]

SUMMARY

A clustered regularly interspaced short palindromic repeats (CRISPR)-like "mimivirus virophage resistance element" (MIMIVIRE) system, which contains specific cascade genes and a CRISPR array against virophages, was reported in mimiviruses. An essential component of the MIMIVIRE system is R354, encoding a nuclease and a likely functional homolog of Cas4. Here we show that R354 is a dual nuclease with both exonuclease and endonuclease activities. Structural analysis revealed that the catalytic core domain of R354 is similar to those of Cas4 and λ exonuclease despite their low sequence identity. R354 forms a homodimer that is important for its exonuclease but not endonuclease activity. Structural comparisons between the active and semi-active states of R354 demonstrated that an activation loop adjacent to the catalytic site is critical for enzymatic activity. Overall, the results suggest that R354 belongs to a novel MIMIVIRE system involved in innate virus immunity and provides a template for the identification of new CRISPR systems in other species.

INTRODUCTION

The clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) system is a unique adaptive immunity mechanism of bacteria and archaea that mediates defense against invasion by foreign microbes (Wiedenheft et al., 2012). Currently, six types of CRISPR-Cas systems (I–VI) have been identified (Hudaiberdiev et al., 2017; Koonin et al., 2017; Makarova et al., 2015). Among them, the type I, II, and V systems consist of several subtype-specific genes (A–D and U of type I; B and C variants of type II, and A, B, and E of type V) (Koonin et al., 2017). These systems also include a variety of Cas proteins, including Cas4, which form complexes that mediate the recognition of foreign DNA sequences (Brouns et al., 2008; Sorek et al., 2013).

Notably, in a recent study, a CRISPR-like mimivirus virophage resistance element (MIMIVIRE) system, which consists of a cascade of genes and a CRISPR array, was reported to protect mimiviruses against virophages (Levasseur et al., 2016) (Figure 1). Specifically, this system functions similarly to the Cas proteins and interferes with virophage invasion (Levasseur et al., 2016) (Figure 1). The nuclease R354, a putatively functional homolog of Cas4 involved in spacer processing for adaption to foreign DNA (Koonin et al., 2017; Plagens et al., 2012; Seed et al., 2013), was found to be essential for MIMIVIRE function.

Although structural information on Cas4 proteins is available, the mechanisms underlying R354-mediated function remain elusive. Here we present the crystal structures of R354 in the active and semi-active states at a resolution of 2.8 Å and 3.0 Å, respectively. The protein is composed of a central channel and a lateral groove, suggesting the entry and exit of the substrate in the cleavage process. Importantly, the crystal structure in the dimerization state, together with biochemical analyses, demonstrates that R354 homodimerization is an exclusive requirement for the exonuclease in double-stranded DNA (dsDNA) but not for the endonuclease activity, revealing a unique regulatory mechanism for R354 function. Altogether, functional and structural analyses suggest that R354 is a functional Cas4-like protein of the viral MIMIVIRE system with a unique regulatory mechanism. Our findings also provide insights into the evolution of nucleases, which might enable the identification of new CRISPR systems in other species.

[1]Division of Respiratory and Critical Care Medicine, State Key Laboratory of Biotherapy, West China Hospital of Sichuan University and Collaborative Innovation Center of Biotherapy, 17th, 3rd Section, Southern Renmin Road, Chengdu, 610041, China

[2]Shanghai Synchrotron Radiation Facility, Zhangjiang Lab, Zhangheng Road 239, Pudong District, Shanghai, 201203, China

[3]Shandong Academy of Medical Sciences, Shandong Institute of Parasitic Disease, Jining, Shandong 272033, China

[4]These authors contributed equally

[5]Lead Contact

*Correspondence: chengwei669@scu.edu.cn

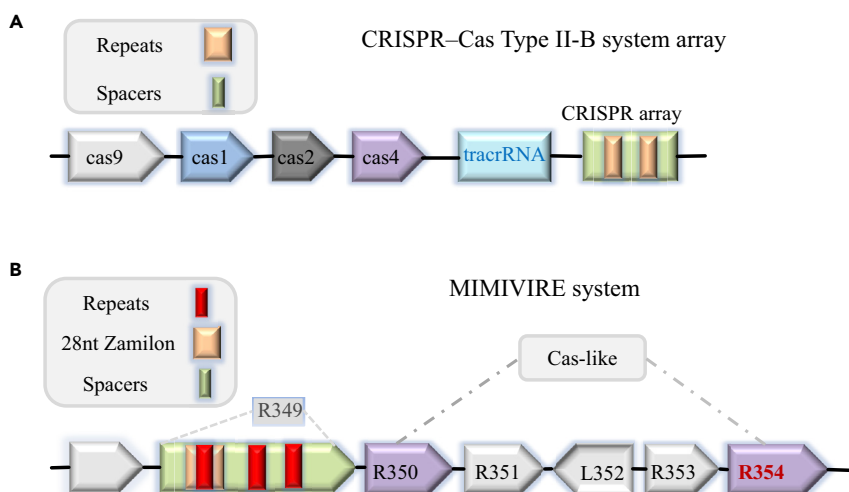https://doi.org/10.1016/j.isci. 2018.04.001

Figure 1. The MIMIVIRE Defense System

(A) Schematic representation showing the prokaryotic CRISPR-Cas system, which is representative of the type II-B system; all components are highlighted in different colors (Hudaiberdiev et al., 2017; Koonin et al., 2017).

(B) Schematic representation showing the viral MIMIVIRE system. Zamilon, a novel virophage, was reported to infect giant viruses from the Mimiviridae family. A specific 28-nucleotide-long sequence in some mimivirus genomes was defined to be identical to the fragment DNA sequence of Zamilon; this pattern is similar to the CRISPR-Cas system, and the unique sequence was named Zamilon. The inserted 28-nucleotide-long sequence is AATCTGATAATGAATCTGATAATGAATC, and the derived 15-nucleotide repeated unit is TGATAATGAATCTGA (Brouns et al., 2008; Gaia et al., 2014).

RESULTS

R354 Is a Dual Nuclease

We investigated the enzymatic activity of R354 using purified protein, as a first step toward the characterization of this MIMIVIRE system protein (Figure S1). We selected plasmid pET21b-LipL as the general substrate for determining R354 endonuclease activity after a digestion-based endonuclease activity assay (Figure 2A). Furthermore, digestion of linear and circular double-stranded plasmid DNA was performed for examining R354 exonuclease activity; the substrate disappeared with time (Figures 2A and 2B). Because most nucleases are divalent metal-dependent enzymes, we performed nuclease assays with different metals. As expected, the nuclease activity of R354 was maximal in the presence of Mg^{2+}, Mn^{2+}, Fe^{2+}, or Zn^{2+}, but not Ca^{2+} (Figure S2), which is consistent with the results of previous reports (Lemak et al., 2013, 2014).

In addition, R354 cleaved a 5′ or 3′ labeled biotinylated dsDNA substrate in the 3′→5′ direction in the exonuclease activity assay (Figure 2C). Conversely, R354 cleaved single-stranded DNA (ssDNA) in both 3′→5′ and 5′→3′ directions (Figure 2D), which is consistent with the function of the Cas4 protein SSO1391 (Lemak et al., 2013). We thus consider the dual nuclease R354 to likely be a functional Cas4 protein in the MIMIVIRE system. Although our data suggest that R354 functions as a dual nuclease, these results are preliminary and further characterization is required to identify the exact mechanism of action.

General Structure of R354

We determined the crystal structures of the wild-type and N-terminal deletion (NTD) mutant of R354 (deletion of amino acids 1–128) at 3.0 Å and 2.8 Å, respectively. Nuclease activity assays showed that the NTD deletion mutant does not affect the activity (Figures 2A and S3). Notably, the structures are also identical; the N-terminal fragment (residues$_{1-137}$) was not observed in the density maps of the wild-type structure. However, because the structure of Del1-128 was determined at higher resolution than that of the wild-type protein, we only discuss the NTD deletion structure for simplicity (Table S1). The protein consists of two subunits (A and B) that form an asymmetric unit (Figure 3B and Table S1). Each subunit is composed of an N-terminus (NT$_{129-194}$), a nuclease region (NUC), a dimerization domain (DD), and a C-terminus (CT; Figure 3A). The NT is connected to the NUC through a loop-linked helix (LH), and two unique linkers, L1 and L2, bridge the NUC and DD and the DD and CT, respectively. The LH-spanning residues 135–190
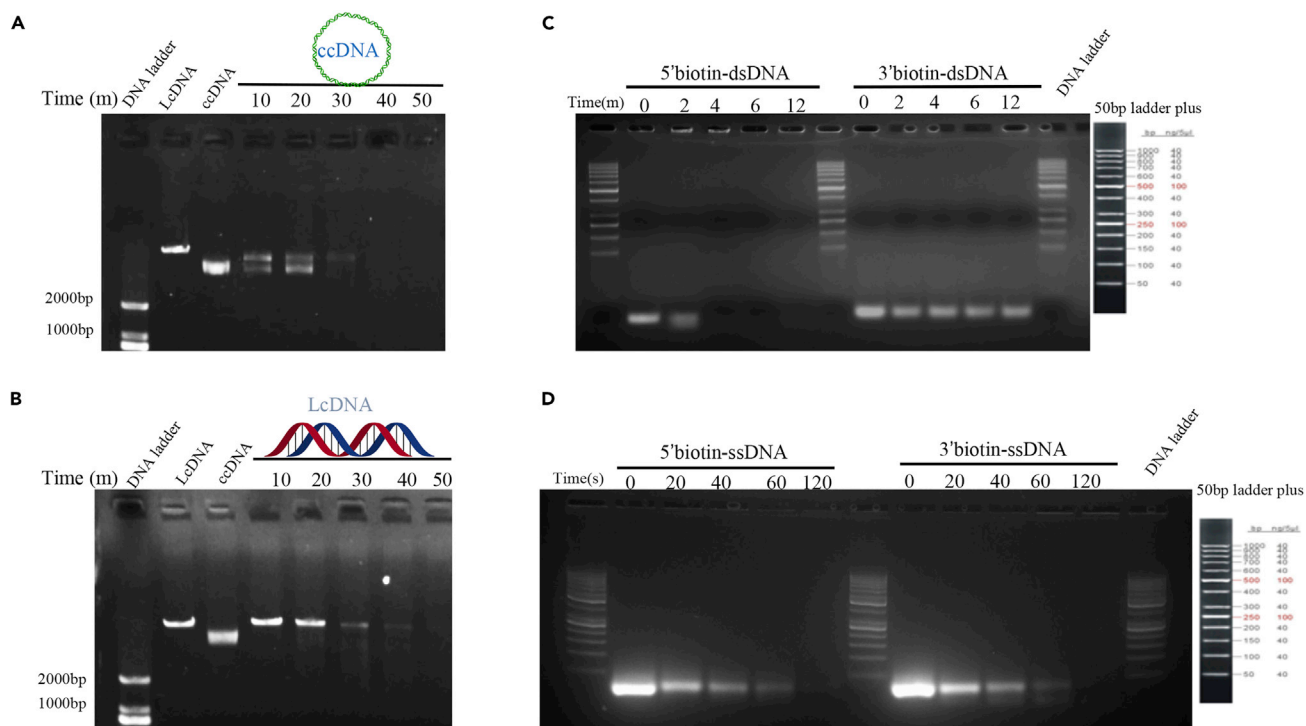
Figure 2. R354 Is a Dual Nuclease

(A) Characterization of the endonuclease activity of R354 by an endonuclease activity assay *in vitro*. ccDNA refers to circular coiled-coil DNA, LcDNA refers to linear coiled-coil DNA.

(B) Characterization of the exonuclease activity of R354 by an exonuclease assay *in vitro*.

(C) Exonuclease activity assay using biotin-labeled dsDNA. R354 was incubated with the labeled dsDNA with blunt ends or 5′ or 3′ overhangs, and the reactions were analyzed by agarose gel electrophoresis (AGE). Results indicate cleavage of the blunt end and the 3′ overhang of the dsDNA.

(D) Exonuclease activity assay using biotin-labeled ssDNA. R354 was incubated with the labeled ssDNA with blunt ends or 5′ or 3′ overhangs, and the reactions were analyzed by AGE. The results indicated cleavage of the blunt ends and the 5′ and 3′ overhangs of the ssDNA. The experiments were repeated three times.

($LH_{135-190}$), which are located at the NT, contain a helix located in the middle of a large loop. $DD_{335-430}$ and $CT_{470-536}$ surround the core structure of $NUC_{192-320}$, which presumably maintains R354 in its functional fold. Overall, the structure of R354 includes a five-stranded β sheet and 16 α helices, which form a central channel and a lateral groove (Figure 3B). The central channel contains a front portal, tunnel, and rear portal, which are possibly used for DNA entry, passage, and exit, respectively (Figure 3C). Several positively charged residues were observed at the front and rear portals of the channel. In contrast, the tunnel harbored several negatively charged residues, suggesting that the protein interacts with polar substrates.

R354 in Active and Semi-active States

Crystallographic analysis of R354 packing revealed that it forms a very tight protein dimer (Figure 4A). Analysis of the dimer interface indicated that protein-protein interactions are mediated by the Lys405 and Glu404 residues of subunits A and B, respectively, which form a strong interaction at the interface (Figure 4B). A clear interaction was also observed between the backbone carbonyl of Met412 of subunit A and the side chain of Tyr430 of subunit B (Figure 4B). In addition, an interaction was formed by Arg427 of subunit A and His415 of subunit B. Notably, other prominent hydrophobic contacts existed among the Leu376, Ile411, and Ile425 residues of each subunit (Figure 4C). These interactions were observed in the structures of both the wild-type protein and the NTD mutant, which demonstrated that the N-terminal 128 residues are not required for dimerization.

Next, we performed biochemical assays to determine whether R354 exists as a dimer in solution. Pull-down and gel filtration assays showed that R354 exists as a dimer (Figures 4D–4F), whereas binding and nuclease activity assays using an R354 variant harboring four amino acid substitutions, namely, Glu404Ala,
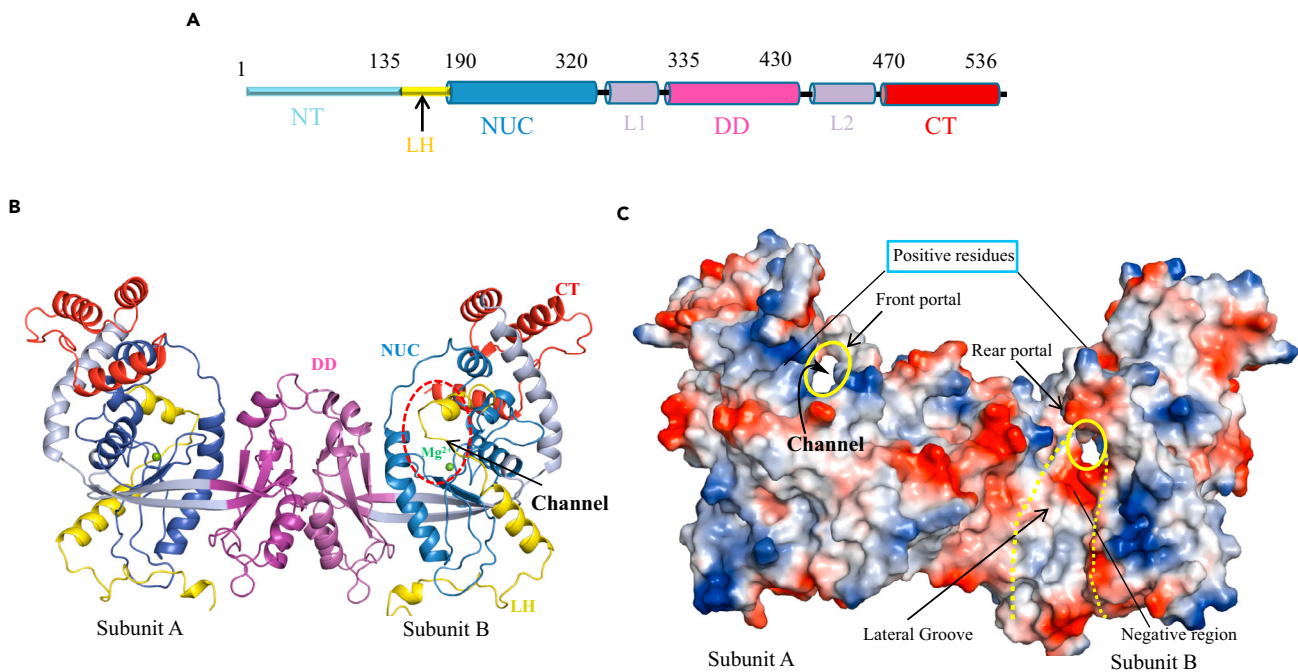
Figure 3. Overall Structure of R354

(A) Graphic depictions of R354 domain organization.

(B) The overall structure of R354; Mg^{2+} is highlighted in green sphere.

(C) Surface representation of R354. The front and rear portals are both highlighted with yellow circles; the lateral groove is indicated by dashed lines.

Lys405Ala, Met412Ala, and Arg427Ala, showed that dimerization was critical for R354 exonuclease activity in both dsDNA and ssDNA (Figures 4E and S4).

Our results show that R354 is a functional homodimer. Furthermore, R354 dimerization provides physical interactions for enhancing the activation of exonuclease function. However, further studies are required to elucidate the mechanism of R354 activation.

We therefore successfully determined the crystal structure of an R354 variant harboring four amino acid substitutions (Glu404Ala, Lys405Ala, Met412Ala, and Arg427Ala) that is in a monomeric state, as supported by pull-down assays (Figure 4D). The monomeric structure is useful for investigating the mechanisms underlying R354 activation (Figure 5A and Table S1). Structural differences between the dimer and monomer were observed (Figure 5B). Specifically, the active site loop (Cys215–His229) was well ordered in the dimer but not observed in the density maps of the monomeric structure. Notably, the tunnel at the active site could not be formed well in the monomer, as shown in Figure 5C. In addition to the conformational changes at the active site, prominent changes were also observed at the dimer interaction interface. Importantly, the large loop (Ser371–Lys389) was flexible in the mutant, and the lengths of certain β sheets and α helices were partially altered (Figure 5B). Moreover, the C-terminal portion of the CT (residues$_{520-538}$) was partially disordered and it did not form the helix observed in the monomeric state (Figures 5B–5D). The importance of this region was verified by enzymatic activity assays (Figure S5). Overall, these observations prompted us to hypothesize that the dimerization state is important for enhancing the activation of R354 exonuclease function.

Based on these conformational observations, we posit that dimerization is required for complete activation of R354. The R354 monomer was loosely folded, highlighting that the rigid fold is required for the full exonuclease activity of R354.

Structural Analysis of the Active Site

A Dali search identified Cas4 protein SSO0001 (Lemak et al., 2013) (PDB ID is 4ic1 with a root-mean-square deviation [RMSD] of 3.4 Å for 162 equivalent Cα atoms), Pcal 0546 (Lemak et al., 2014) (PDB ID is 4r5q with an
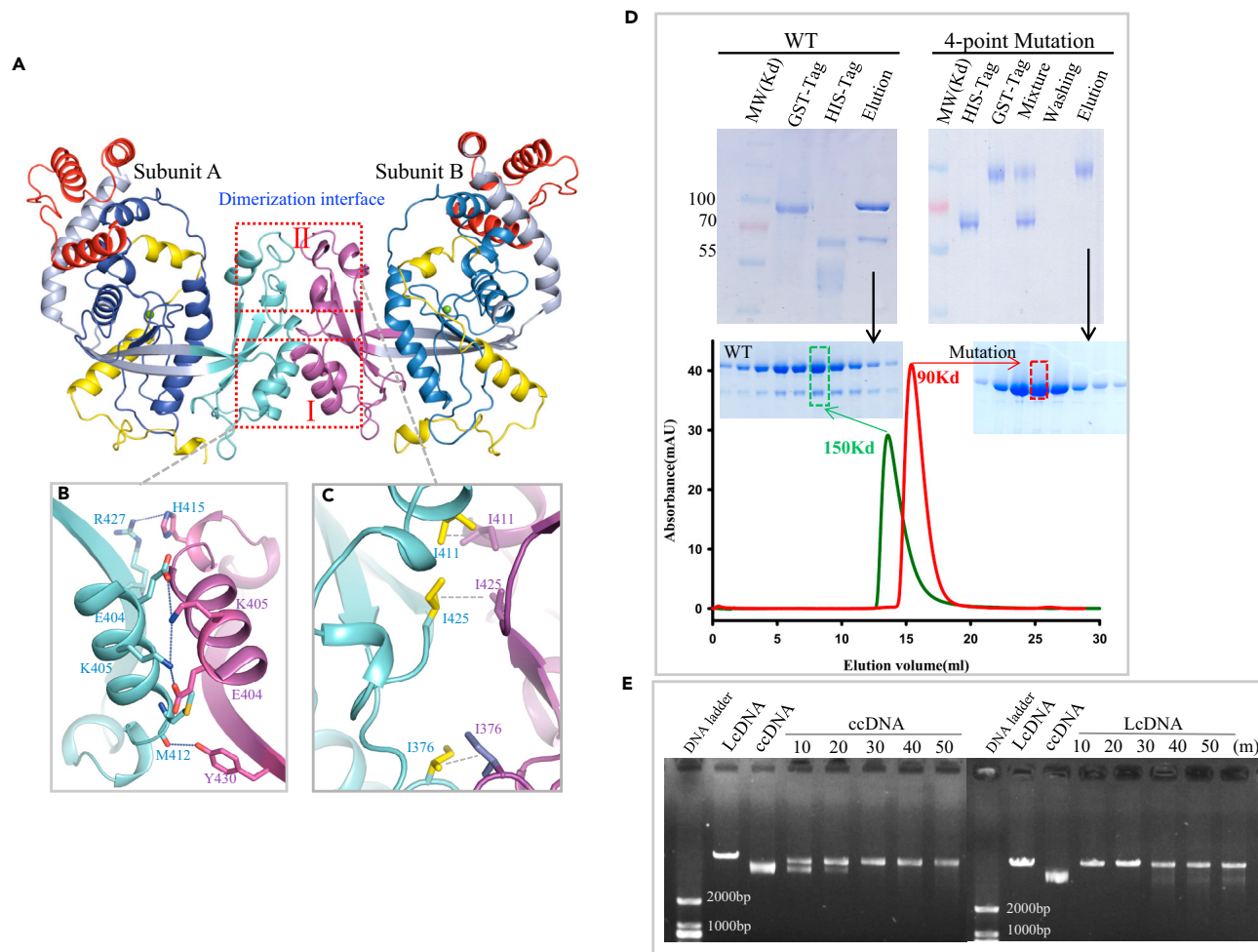
Figure 4. Dimerization of R354

(A) Dimerization surface I and surface II formed by the interaction of subunit A and subunit B of R354.

(B) Close up of the interactions of interface I. Hydrogen bonds and ionic bonds are indicated by blue dashed lines.

(C) Close up of the interactions of interface II. The hydrophobic interactions are indicated by gray dashed lines.

(D) Verification of dimerization of R354 in solution by pull-down assay and gel filtration analyses; details described in Transparent Methods.

(E) Verification of the dimer-dependent exonuclease activity of R354 using a variant harboring four point mutations (E404A, K405A, M412A, and R427A). LcDNA refers to linear coiled-coil DNA; ccDNA refers to circular coiled-coil DNA.

RMSD of 4.2 Å for 143 equivalent Cα atoms), and λ exonuclease (Kovall and Matthews, 1997; Zhang et al., 2011) (PDB ID is 3sm4 with an RMSD of 2.8 Å for 186 equivalent Cα atoms) as being structurally similar to R354. This structural similarity was largely attributed to the nuclease domain (NC$_{186\_346}$) despite low levels of sequence identity (Figures 6A and S6). Together, these findings indicate that the CRISPR system could have evolved in species other than bacteria and archaea, including viruses. Moreover, the findings suggest that the nuclease and CRISPR systems of bacteria and mimiviruses evolved from a cenancestor and that the R354 coding sequence was acquired via vertical inheritance (Figure S7). Therefore we can interpret the catalytic mechanism of R354 by structural comparison.

In contrast to Cas4 and λ exonucleases, which require a toroidal state for their exonuclease activities, dimerization was essential for the exonuclease but not the endonuclease activity of R354. These findings suggest that R354 may employ a unique regulatory mechanism. However, the interactions between Mg^{2+} and residues Asp268, Glu289, and Lys291 within the active site of R354 were similar to those of Mn^{2+}, Asp99, Glu113, and Lys115 in Cas4 SSO0001 and Mg^{2+}, Asp123, Glu136, and Lys138 in Cas4 Pcal 0546 (Figure 6B). These three active site residues are critical for protein function, and a single mutation
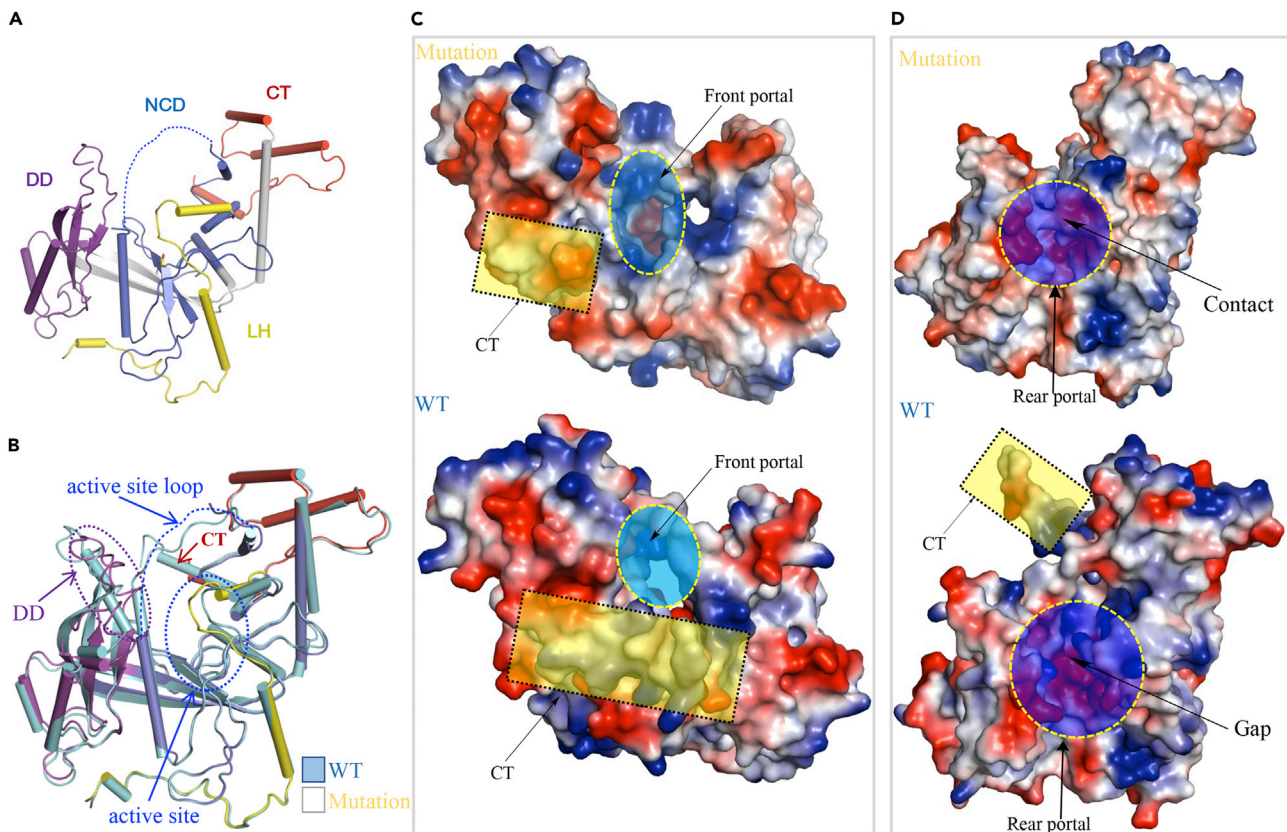
Figure 5. The Semi-active State of R354

(A) Overall structure of R354 harboring the four point mutations (E404A, K405A, M412A, and R427A). The active site loop region indicated by the dotted line, which is not observed in the electron density map and the electron density map for the C terminus, is also not observed.

(B) Structural comparisons of the active and semi-active states of R354. The active state is more rigid than the semi-active state, particularly in the portions DD, NCD, and CT.

(C) The electrostatic representations viewed from the front portal for both semi-active and active states, the different front portals, and the CT portions are highlighted in blue and yellow colors, respectively.

(D) The electrostatic representations viewed from the rear portal for both semi-active and active states. The different rear portals and the CT portions are highlighted in green and yellow colors, respectively.

can almost abolish exonuclease activity. In addition, other critical residues, including Gln316, Tyr313, and Tyr312, were also similar to Glu131, Tyr135, and Tyr148 in Cas4 SSO0001 and Glu149, Tyr153, and Tyr166 in Cas4 Pcal 0546 (Figure 6B); these three residues are likely associated with substrate passage through the central channel. The Mg^{2+} in the catalytic site possibly enhances the catalytic rate, although the detailed mechanism is not fully understood. Possibly, metal ions coordinate the oxygen atom to trigger a nucleophilic attack and form the leaving group during cleavage of the scissile phosphodiester bond (Horton and Perona, 2001; Steitz and Steitz, 1993). The structural and functional similarities and the presence of conserved residues thus suggest that R354 is a Cas4-like protein.

To reveal the interaction of DNA with R354, we superimposed the structures of λ exonuclease in complex with dsDNA (Zhang et al., 2011) (PDB ID: 3SM4) on our R354 structure, resulting in a good fit between DNA from the λ exonuclease structure and the surface of R354 (Figure 6C). The modeling indicated obviously that R354 has two groups of residues for DNA binding and digesting, which are located at the front portal and inner tunnel (Figures 3C and 6C), respectively; some other residues adjacent to the active site are possibly involved in DNA moving. To further verify the importance of these residues, we mutated the residues within the central channel, including those at the entry and exit sites, to alanine and performed enzymatic activity assays with the mutants. The residues at the front portal site are probably involved in contacting the downstream portion of the dsDNA for entering the tunnel (Arg205, Arg296, and Tyr313), residues at
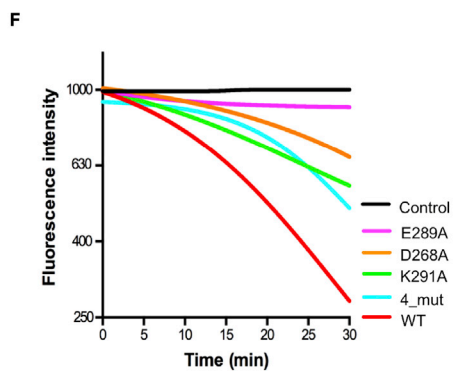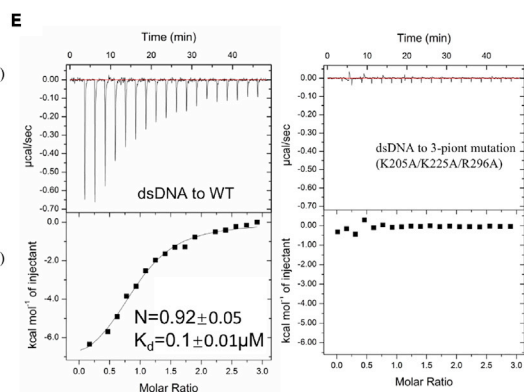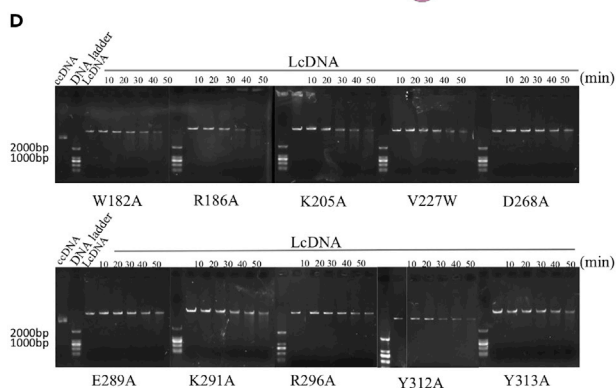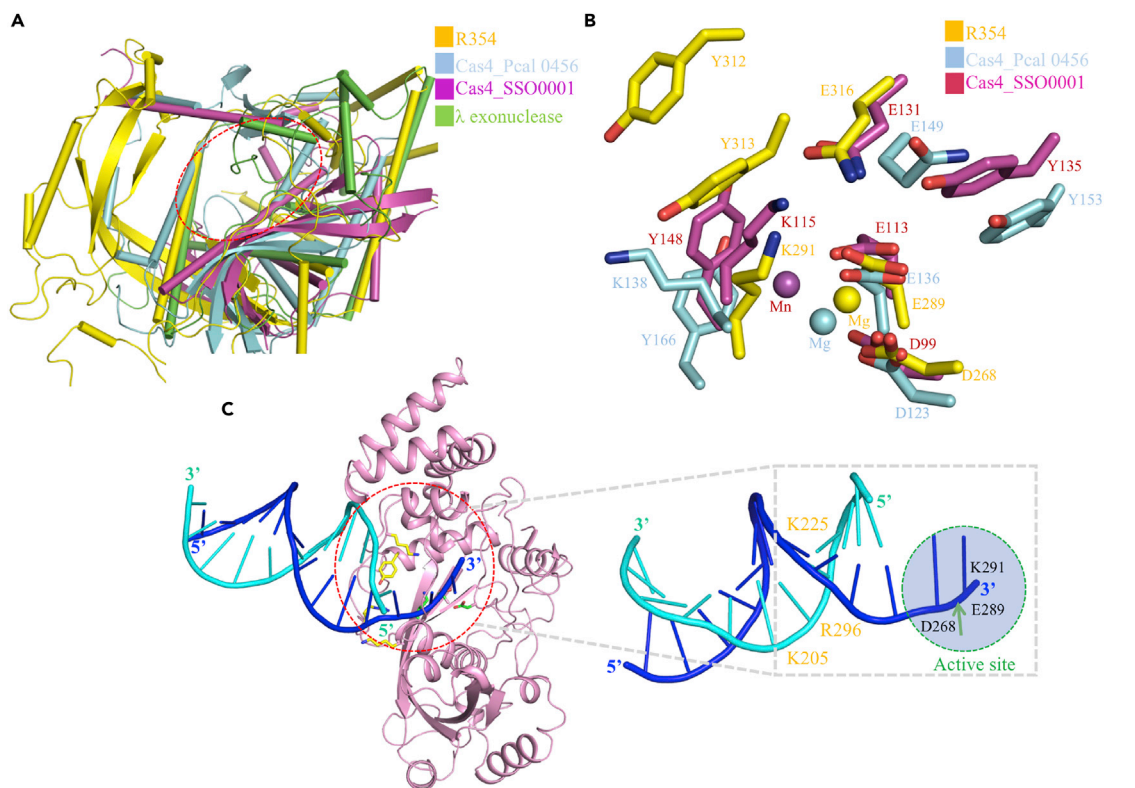
Figure 6. Structural Analysis

(A) Structural comparisons of R354, Cas4_Pcal 0456, Cas4_SSO0001, and λ exonuclease. The similarities between the core catalytic domains of the four proteins are highlighted by red dashed lines.

(B) Comparison of active sites of R354 and Cas4; some residues are conservative in both active sites, which suggests that R354 has similar function to Cas4.

(C) A cartoon presentation of the dsDNA modeled into the structure of R354 (blue for the cleaved strand with the 3′-OH and cyan for the other strand), which shows interactions between DNA and R354. DNA was modeled by superimposing the catalytic core domain of R354 and λ exonuclease (PDB ID 3SM4 Zhang et al., 2011). The red dotted ring is interpreted in the right panel. Residues putatively interacting with the DNA substrate and involved in catalytic reaction are shown as sticks in the left panel. The cut site of DNA strand and the active site of R354 are indicated with highlighted label; the key residues are labeled in the right panel.

(D) Verification of key residues involved in R354 enzymatic activity using activity assays. LcDNA refers to linear coiled-coil DNA; ccDNA refers to circular coiled-coil DNA.

(E) Isothermal titration calorimetry (ITC) measurement of dsDNA binding. Representative ITC raw data and binding trace for dsDNA titrated into the solution containing wild-type (WT) R354 in the left panel: $K_d = 0.1 \pm 0.01\ \mu M$; $N = 0.92 \pm 0.05$. The variant harboring three point mutations (K205A/K22A/R296A) can disrupt the dsDNA binding in the right panel.

(F) Quantitative exonuclease assays. (Left) Comparison of exonuclease activities of wild-type and representative variants of R354. The reactions began to record (m = 0) at the signal stabilized by the addition of enzyme to reaction buffer. (Right) The reaction parameters are shown in the table. The first column indicates the rate determined by the initial slope of the reaction. The percent of DNA substrate digested at 30 min is shown in the second column by comparison of detected fluorescence relative to reactions between positive (WT) and negative controls. Data are measured with three independent replicates for standard errors.

the hydrophobic wedge may be required for DNA movement (Val227), the active site in the central channel is required for catalysis (Asp268, Glu289, and Lys291), and the residues at the rear portal site and slide groove are possibly used for phosphate exit and sliding out of the processed ssDNA (His145, Arg171, Gln177, Trp182, and Arg186). Substitution of each of the above residues adjacent to the active site with alanines (Trp182Ala, Arg186Ala, Lys205Ala, Val227Ala, Asp268Ala, Glu289Ala, Lys291Ala, Arg296Ala, Tyr312Ala, and Tyr313Ala) nearly abolished the exonuclease activity of R354 (Figure 6D), highlighting their functional significance. In addition, isothermal titration calorimetry (ITC) measurements of some variants (Arg205Ala, Arg296Ala, and Tyr313Ala) supported the proposed binding sites for dsDNA (Figure 6E). Moreover, a quantitative exonuclease assay was performed for the representative residues with the corresponding mutants using fluorescence experiments (Figure 6F). Only the mutant E289A, with mutations at key residues of the active site (Asp268, Glu289, and Lys291), had almost complete abolishment of exonuclease activity. The four-residue mutant (Glu404Ala, Lys405Ala, Met412Ala, and Arg427Ala) showed a marked decrease in exonuclease activity, which demonstrated that R354 in the dimer state is essential for its full enzymatic activity.

DISCUSSION

Here we report the first structure of the Cas4-like protein, R354, in the MIMIVIRE system (Figure 1), which is involved in defending against virophage invasion. Nuclease activity assays supported that R354 is probably a dual nuclease (Figures 2, 6F, and S2), and the exonuclease activity was verified by a quantitative assay (Figure 6F); however, the more explicit dual mechanism of the enzyme needs more evidence. Coupled with biochemical data, our structural studies provide insights into the unique mechanisms of R354 exonuclease activity that are enhanced by the dimerized state (Figure 6F), but not the endonuclease activity, which revealed the unique mechanism of R354. Notably, the different states of R354 indicate that the mechanism of R354 activation is distinct from those of other nuclease family members.

Considering that Cas4 SSO0001 (Lemak et al., 2013), Pcal 0546 (Lemak et al., 2014), and λ exonuclease (Kovall and Matthews, 1997; Zhang et al., 2011) all form toroidal oligomers, whereas R354 forms a dimer, we speculated that R354 exhibits a sequential catalytic mechanism different from those of its homologs. Indeed, the major difference between these molecules was that activated R354 dimers could be distinguished from the toroidal oligomers, suggesting that R354 possibly catalyzes one strand of dsDNA with the dimer conformation, resulting in the release of the other strand from the central channel of the subunit; the released strand then enters the other subunit for cleavage. Therefore the mechanism of R354 activation is different from those of the toroidal oligomers. A hypothetical catalytic mechanism of R354 is shown in Figure 7. The 3′-OH group of the dsDNA first binds to residues Lys205, Lys225, and Arg296, primarily via polar interactions. Moreover, hydrophobic contacts, particularly with Val227, drive the dsDNA close to the catalytic channel, and the putative dsDNA binding site was supported by ITC measurement (Figure 6E). However, the mechanisms of reaction catalysis and energy generation for the reaction are not understood. An enzyme-DNA-metal complex is formed when the dsDNA enters the active site; the divalent metal
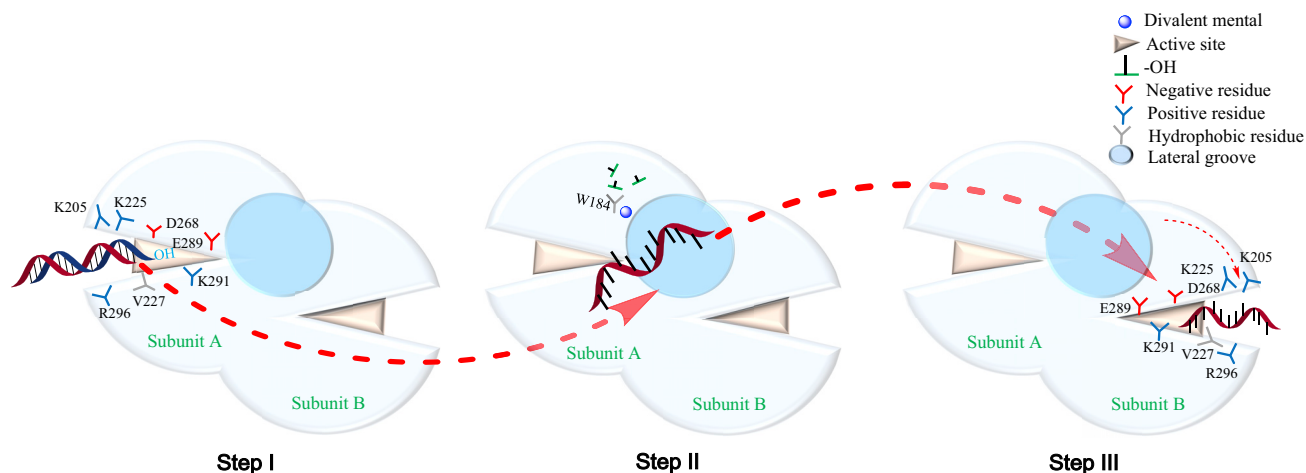
Figure 7. A Working Model of R354

Cartoon depicting the working model for R354 exonuclease catalytic reaction. In step I, the 3'-OH group of the dsDNA first binds to residues Lys205, Lys225, and Arg296, primarily via polar interactions. Moreover, hydrophobic contacts, particularly with Val227, drive the dsDNA close to the catalytic channel. In step II, one strand is digested into single nucleotides and the other strand slides out from the rear portal. In step III, the sliding strand is digested by subunit B without directional priority.

activates the scissile phosphate along with catalytic residues Asp268, Glu289, and Lys291, and residues W182 and R186 capture the leaving phosphate groups (Figures 6C–6F). One strand of dsDNA is then cleaved, whereas the other strand likely passes through the tunnel and remains bound to the lateral open groove (Figure 1D). This strand then slides into the other tunnel for cleavage, which may explain why R354 exhibits more efficient catalytic activity than the λ exonuclease.

In summary, we report the first structure of R354, an essential viral enzyme in the novel MIMIVIRE CRISPR-Cas-like system. The structure revealed a unique mechanism of R354 activation, which exhibits dual nuclease activity. Moreover, our findings suggest that R354 is a Cas4 homolog, and therefore it extends the list of novel MIMIVIRE spacers involved in innate immunity against virophages. In addition, this protein provides a template for the identification of new CRISPR-Cas systems in other species. However, the function of R354 similar to Cas4, the component of the CRISPR system, is still unclear due to the limitations of the current study. Future work is required to determine how R354 contributes to the MIMIVIRE system function for viral immunity.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND SOFTWARE AVAILABILITY

The atomic coordinates and structure factors of the reported crystal structures have been deposited in the Protein Data Bank (PDB), under the following accession codes: 5YET is for wild-type R354; 5YEU is for four-point mutation R354.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Transparent Methods, seven figures, and one table and can be found with this article online at https://doi.org/10.1016/j.isci.2018.04.001.

ACKNOWLEDGMENTS

REFERENCES

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and Van, J.D.O. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960.

Gaia, M., Benamar, S., Boughalmi, M., Pagnier, I., Croce, O., Colson, P., Raoult, D., and La Scola, B. (2014). Zamilon, a novel virophage with Mimiviridae host specificity. PLoS One *9*, e94923.

Horton, N.C., and Perona, J.J. (2001). Making the most of metal ions. Nat. Struct. Biol. *8*, 290–293.

Hudaiberdiev, S., Shmakov, S., Wolf, Y.I., Terns, M.P., Makarova, K.S., and Koonin, E.V. (2017). Phylogenomics of Cas4 family nucleases. BMC Evol. Biol. *17*, 232.

Koonin, E.V., Makarova, K.S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. *37*, 67.

Kovall, R., and Matthews, B.W. (1997). Toroidal structure of λ-exonuclease. Science *277*, 1824–1827.

Lemak, S., Beloglazova, N., Nocek, B., Skarina, T., Flick, R., Brown, G., Popovic, A., Joachimiak, A., Savchenko, A., and Yakunin, A.F. (2013). Toroidal structure and DNA cleavage by the CRISPR-associated [4Fe-4S]-cluster containing Cas4 nuclease SSO0001 from *Sulfolobus solfataricus*. J. Am. Chem. Soc. *135*, 17476–17487.

Lemak, S., Nocek, B., Beloglazova, N., Skarina, T., Flick, R., Brown, G., Joachimiak, A., Savchenko, A., and Yakunin, A.F. (2014). The CRISPR-associated Cas4 protein Pcal_0546 from *Pyrobaculum calidifontis* contains a [2Fe-2S] cluster: crystal structure and nuclease activity. Nucleic Acids Res. *42*, 11144–11155.

Levasseur, A., Bekliz, M., Chabrière, E., Pontarotti, P., Scola, B.L., and Raoult, D. (2016). MIMIVIRE is a defence system in mimivirus that confers resistance to virophage. Nature *531*, 249–252.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., and Haft, D.H. (2015). An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. *13*, 722–736.

Plagens, A., Tjaden, B., Hagemann, A., Randau, L., and Hensel, R. (2012). Characterization of the CRISPR/Cas subtype I-A system of the

hyperthermophilic crenarchaeon *Thermoproteus tenax*. J. Bacteriol. *194*, 2491–2500.

Seed, K.D., Lazinski, D.W., Calderwood, S.B., and Camilli, A. (2013). A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature *494*, 489–491.

Sorek, R., Lawrence, C.M., and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu. Rev. Biochem. *82*, 237–266.

Steitz, T.A., and Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. Proc. Natl. Acad. Sci. USA *90*, 6498–6502.

Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature *482*, 331–338.

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

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
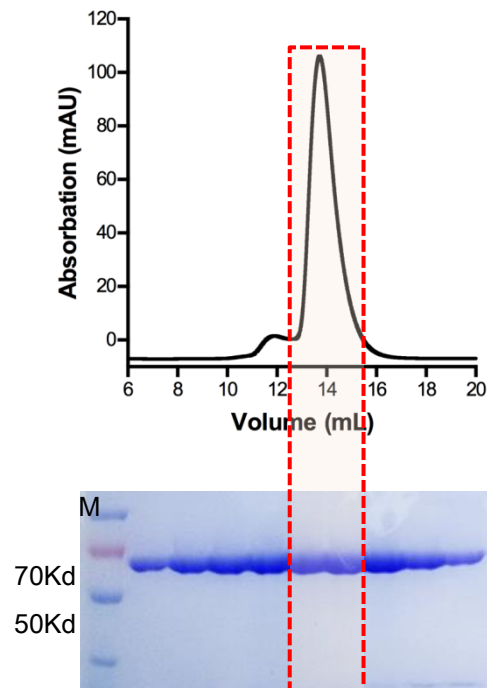

Figure S1. High quality protein purification of R354, related to Figure 2.

The upper panel indicates protein purification by gel filtration. The lower panel shows

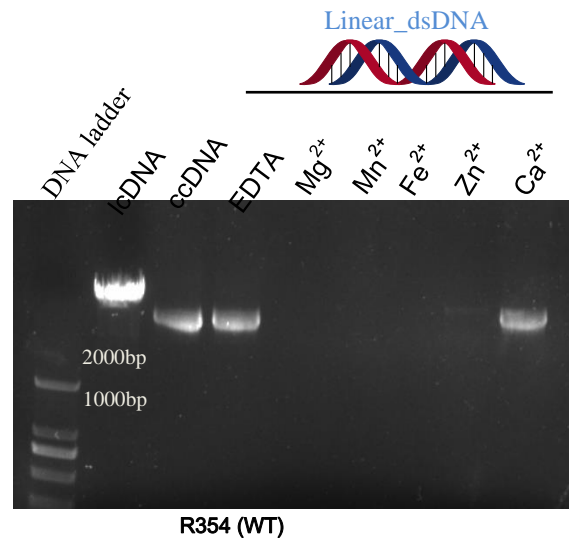SDS-PAGE detection by Coomassie Brilliant Blue staining

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.*
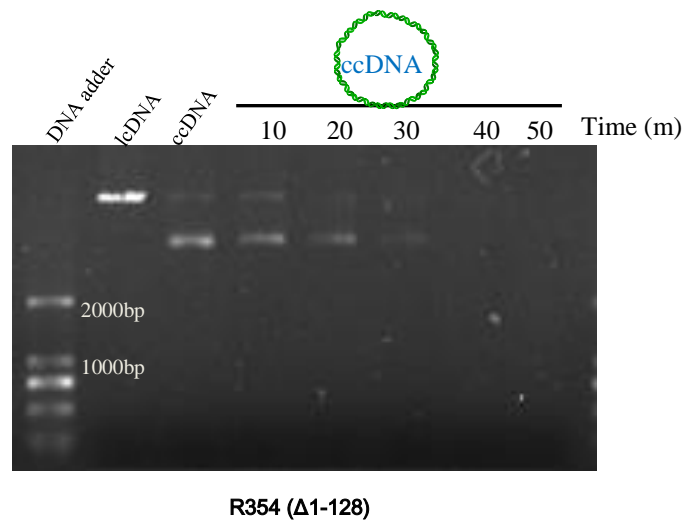
Figure S3

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

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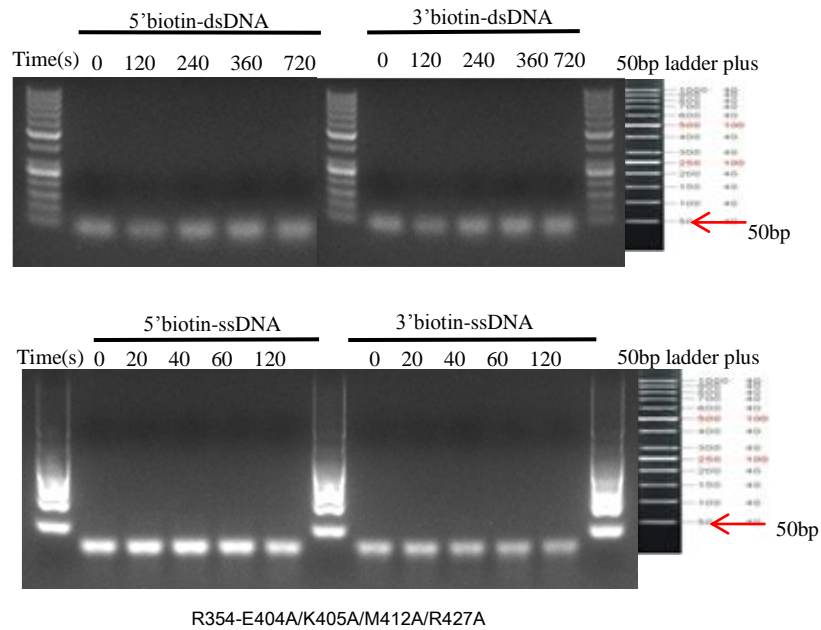
Figure S4. Exo-nuclease activity of R354-E404A/K405A/M412A/R427A, related to Figure 4.

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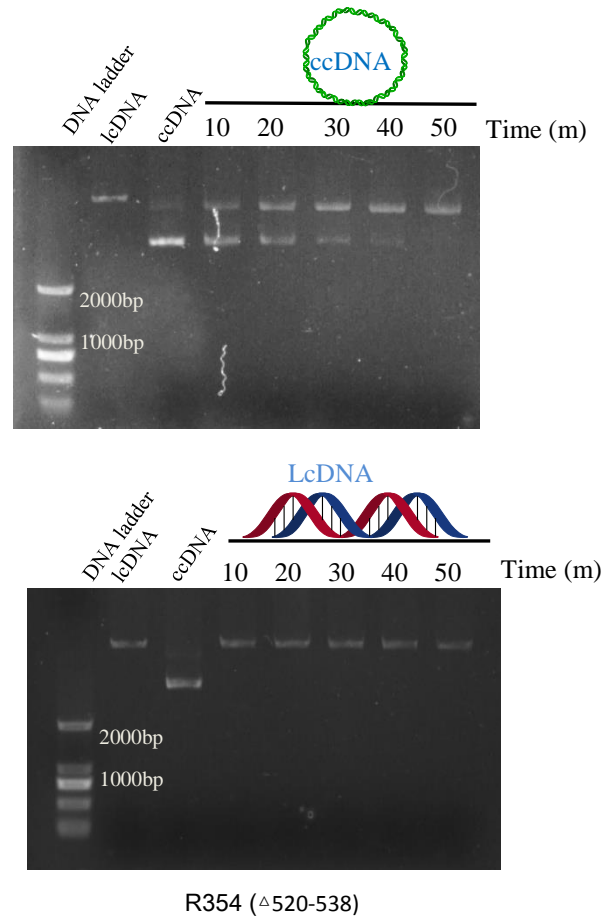
Figure S5. Nuclease activity assay for the R354 C-terminal-deletion variant, related to Figure 5. CcDNA refers to circular coiled-coil DNA; LcDNA refers to linear coiled-coil DNA.

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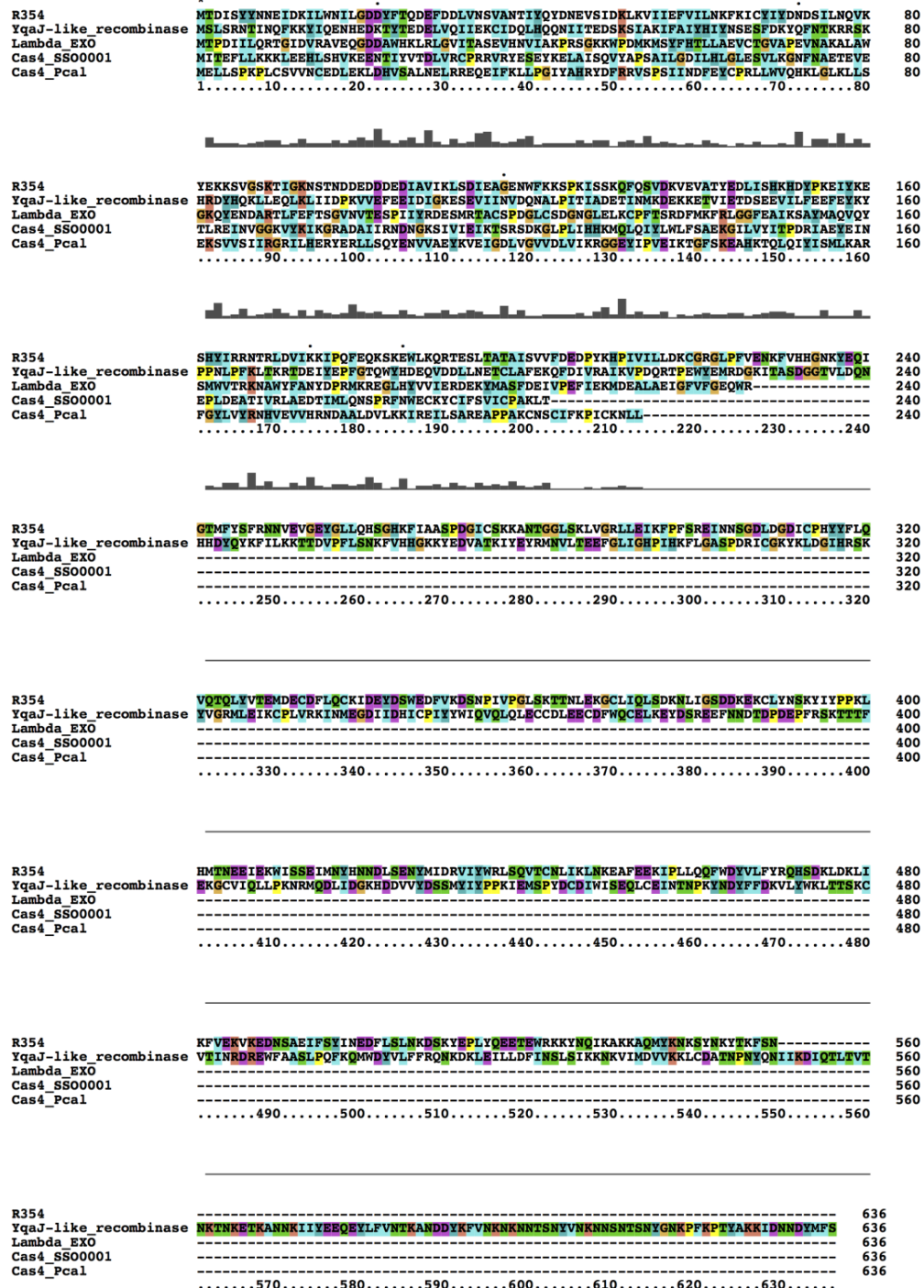
Figure S6. Amino acid sequence alignment of R354, Cas4, and λ exonulease, related to Figure 6.

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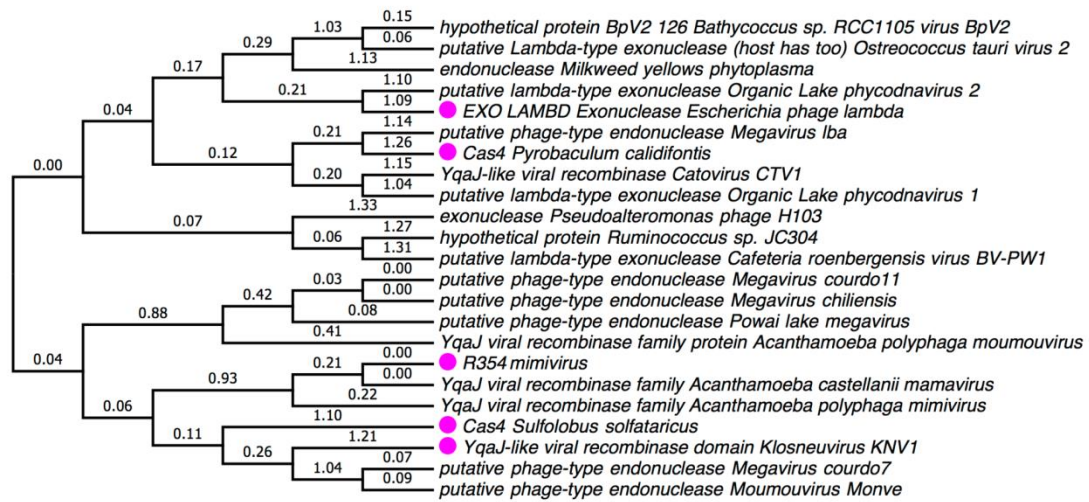
Figure S7. Phylogenetic tree distribution of representative members of the nuclease family, related to Figure 6.

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 (Δ 1-128) PDB ID: 5YET	SeMet-R354 (Δ 1-128)	Four-point mutations R354 (Δ 1-128) 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.