

Staphylococcus epidermidis Csm1 is a 3'–5' exonuclease

Nancy F. Ramia¹, Li Tang¹, Alexis I. Coczaki¹ and Hong Li^{1,2,*}

¹Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA and ²Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306, USA

Received May 22, 2013; Revised September 17, 2013; Accepted September 18, 2013

ABSTRACT

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) offer an adaptive immune system that protects bacteria and archaea from nucleic acid invaders through an RNA-mediated nucleic acid cleavage mechanism. Our knowledge of nucleic acid cleavage mechanisms is limited to three examples of widely different ribonucleoprotein particles that target either DNA or RNA. Staphylococcus epidermidis belongs to the Type III-A CRISPR system and has been shown to interfere with invading DNA *in vivo*. The Type III-A CRISPR system is characterized by the presence of Csm1, a member of Cas10 family of proteins, that has a permuted histidine–aspartate domain and a nucleotidyl cyclase-like domain, both of which contain sequence features characteristic of nucleases. In this work, we show *in vitro* that a recombinant *S. epidermidis* Csm1 cleaves single-stranded DNA and RNA exonucleolytically in the 3'–5' direction. We further showed that both cleavage activities are divalent-metal-dependent and reside in the GGDD motif of the cyclase-like domain. Our data suggest that Csm1 may work in the context of an effector complex to degrade invading DNA and participate in CRISPR RNA maturation.

INTRODUCTION

Bacteria and archaea are greatly outnumbered by bacteriophages and other mobile genetic elements such as plasmids (1–5). Accordingly, these organisms have developed different strategies to protect themselves against the infections (6). Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) represent a recently identified prokaryotic defense system that resembles RNA interference in eukaryotes (7). The CRISPR-based

defense systems use small RNA, termed CRISPR (cr)RNA, to guide the destruction of foreign RNA or DNA (7–11). A CRISPR locus is typically composed of a leader sequence, several CRISPR-associated (*cas*) genes and an array of repeat sequences of 30–50 nucleotides separated by distinct spacers of similar sizes. The sequences of some spacers were found to match those from invading phages or plasmids (12–15), which offered the initial evidence for the RNA interference defense mechanism.

There are three functional stages of the CRISPR-based immunity. These include integration of new DNA fragments from the invader, transcription and processing of crRNA and degradation of the invading DNA. The *cas* gene-encoded Cas proteins play critical roles in each of these functional stages. Cas proteins, other than Cas1 and Cas2 that are predicted to function in integration of new spacers, vary largely in sequence and motifs (16–18), which makes it difficult to predict and elucidate their functional roles. The CRISPR–Cas systems are classified into three major types (I, II and III) and their respective subtypes based on the presence and phylogeny of signature *cas* genes encoding crRNA processing and effector complex proteins (18). Both Type I and III systems use a single endoribonuclease to generate functional crRNA that comprise an 8-nt 5' tag derived from the 3' region of the repeat and a spacer sequence, and in some cases, an additional 3' tail derived from the 5' region of the following repeat (9,19–22). Each mature crRNA is loaded onto a multi-subunit effector complex made of Cas proteins that then cleaves RNA or DNA complementary to the crRNA (9,11,20,23–25). Different from Type I and III systems, Type II systems require *trans*-activating CRISPR RNA (tracrRNA) along with the host-encoded RNase III to generate crRNA and a single signature protein bound with tracrRNA and crRNA to degrade foreign DNA (26–28). The effector complexes studied so far differ dramatically in composition and the mode of nucleic acid cleavage. Examples of both Types I and II effector complexes have been shown to target invading DNA, although by very different mechanisms (24,27–29).

*To whom correspondence should be addressed. Tel: +1 850 644 6785; Fax: +1 850 644 7244; Email: hong.li@fsu.edu

In contrast, examples of the Type III effector complexes have been shown to target either RNA or DNA (10,11,25).

Type I effector complexes are characterized by the presence of Cas3, a protein possessing the helicase (DExH) domain, and a ribonucleoprotein complex containing the Cas8 superfamily of proteins (17,18). While the Cas8 family of proteins is believed to play an important role in the effector complex assembly (30,31), Cas3 is the key player in DNA degradation (24). Cas3 often has a fused nuclease histidine–aspartate (HD) domain, but in some organisms, works with a separately encoded HD protein (7,16,18,32). Multiple reports have shown a metal-dependent nuclease activity in the HD domain of Cas3 or the isolated HD protein (33–35). The best studied Type I effector complex is the Cascade (for CRISPR-associated complex for antiviral defense) of *Escherichia coli*. Cascade consists of five Cas proteins and a crRNA and binds invading DNA complementary to the crRNA. Structural and biochemical studies have provided insights on the functional mechanism of the Cascade complex (9,24,36,37). Following base-pairing of crRNA with the complementary DNA strand, the non-complementary DNA strand is displaced, leading to formation of an R-loop (24,36,37). The R-loop formation recruits Cas3 (23,24) that uses its nuclease domain to nick the target DNA and its helicase domain to unwind DNA for further degradation of the negative strand (33,35,38).

Type II effector complexes are characterized by the presence of a multi-domain protein Cas9 that possesses an N-terminal RuvC and an HNH nuclease domain located in the middle of the protein (7). Cas9 bound with crRNA and tracrRNA is sufficient for silencing of the invading DNA (27,28). Cas9 produces double-strand breaks by cleaving the DNA strand complementary to crRNA via its HNH domain, and the non-complementary strand via its RuvC-like domain (28), possibly also through the formation of an R-loop structure (27,28,39). Type II systems comprise the simplest CRISPR effector complexes with multiple potential applications for genetic engineering and DNA modeling (28,40,41).

Type III effector complexes contain a protein of the Cas10 superfamily and a number of repeat-associated mysterious proteins (16–18). Cas10 also possesses the HD nuclease domain, although with a permuted arrangement of catalytic residues (7), and a nucleotidyl cyclase-like domain (17,42,43). Subtype III-B effector complex, called the Cmr complex, targets invading RNA for destruction (11,25). Active Cmr complex can be reconstituted *in vitro* from its protein subunits (Cmr1–Cmr7) and a crRNA (11,25,43,44). However, exhaustive search for the catalytic subunit has not yet led to a positive identification. Despite prominent features similar to those found in nucleases, neither the permuted HD domain nor the cyclase-like domain of the characteristic protein, Cas10 (Cmr2), is found to be directly involved in cleaving the crRNA-bound target RNA (43).

Interestingly, another studied Type III system belonging to subtype III-A appears to target invading DNA (10,45). The CRISPR system of the clinically isolated

Staphylococcus epidermidis strain RP62A contains Type III-A proteins Csm1–Csm6 (Figure 1) where the Cas10 protein, Csm1, has a similar domain arrangement as that in Cmr2. In *S. epidermidis* RP62A, a DNA plasmid containing a CRISPR target interrupted by a self-splicing intron sequence was found to be evaded by its CRISPR system, suggesting that the mRNA was not targeted (10). Like Cmr2, Csm1 contains both the permuted HD and the nucleotidyl cyclase-like domains (7,17) and has been hypothesized to be responsible for cleaving DNA (18). However, no biochemical characterization of this activity is available, and the mechanism of interference for the Type III-A is still poorly understood.

In this work, we report characterization of a nuclease activity of *S. epidermidis* Csm1 (SeCsm1). Our results show that SeCsm1 cleaves single-stranded (ss)DNA and RNA in a divalent metal-dependent fashion, and exhibits a 3′–5′ exonuclease activity. Mutational analysis suggests that the catalytic center resides in the cyclase-like rather than the HD domain, although the HD domain also plays a functional role in DNA cleavage by SeCsm1. Our results provide supporting evidence that the Type III-A effector is a DNA-targeting complex.

MATERIALS AND METHODS

Cloning and expression of Serp2461

Csm1 (SERP2461) was amplified from genomic DNA and cloned into two Gateway vectors (Invitrogen) to create pDest-SERP2461 and Pvp16-SERP2461 that express either a polyhistidine-tagged (His-SeCsm1) or a Maltose Binding Protein-fused protein (MBP-SeCsm1). The MBP-SeCsm1 contains a TEV cleavage site that allows removal of the MBP fusion tag (Figure 2A). To express either Csm1 protein, 2–6 l of Luria–Bertani medium was inoculated with a saturated overnight culture of *E. coli* cells (Rosetta DE3) transformed with the plasmid and grown at 37°C. Protein expression was induced by addition of 0.5 mM of isopropyl-β-thiogalactoside at OD 0.6–0.8. After growth for additional 16 h at 18°C, the cells were harvested by centrifugation.

Purification of MBP-SeCsm1

While the pDest-SERP2461 clone produced only insoluble His-SeCsm1 protein, Pvp16-SERP2461 resulted in soluble MBP-SeCsm1. To purify the MBP-SeCsm1 protein, cells were lysed using a homogenizer in ice-cold buffer containing 25 mM sodium phosphate pH 7.4, 400 mM NaCl, 10% glycerol and 5 mM of dithiothreitol, and the protease inhibitor 0.1% v/v phenylmethylsulfonyl fluoride. The cell lysate was clarified by centrifugation at 22000 rpm for 30 min at 4°C and the supernatant was loaded onto an amylose column pre-equilibrated with the lysis buffer. The column was then washed with the lysis buffer plus 1 mM maltose before elution with the lysis buffer containing additional 50 mM of maltose. The eluted protein was further purified by monoQ ion exchange and size-exclusion chromatography on a S200 16/60 Superdex column. The final protein was concentrated to 5 mg/ml and stored in a buffer

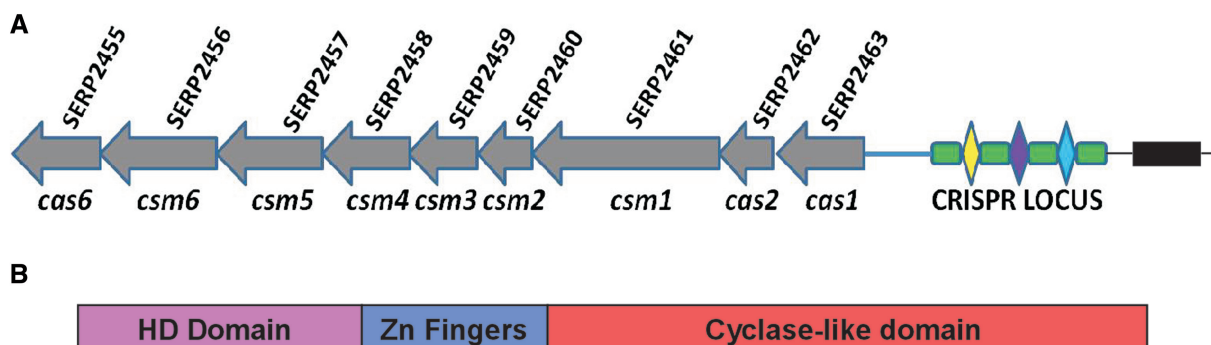


Figure 1. Schematic organization of the CRISPR locus of *Staphylococcus epidermidis* RP62a and of Csm1 protein domains. (A) Repeats are represented as green boxes and spacers as colored blocks. The leader region is shown as a black box and the CRISPR-associated (cas) genes are shown as block arrows with their locus names according to the comprehensive microbial resource labeled on top and common names on the bottom. (B) The graphical representation of the Csm1 protein domains. The histidine–aspartate domain (HD) is shown in light blue, Zn fingers domain is in pink and the nucleotidyl cyclase-like domain is shown in orange.

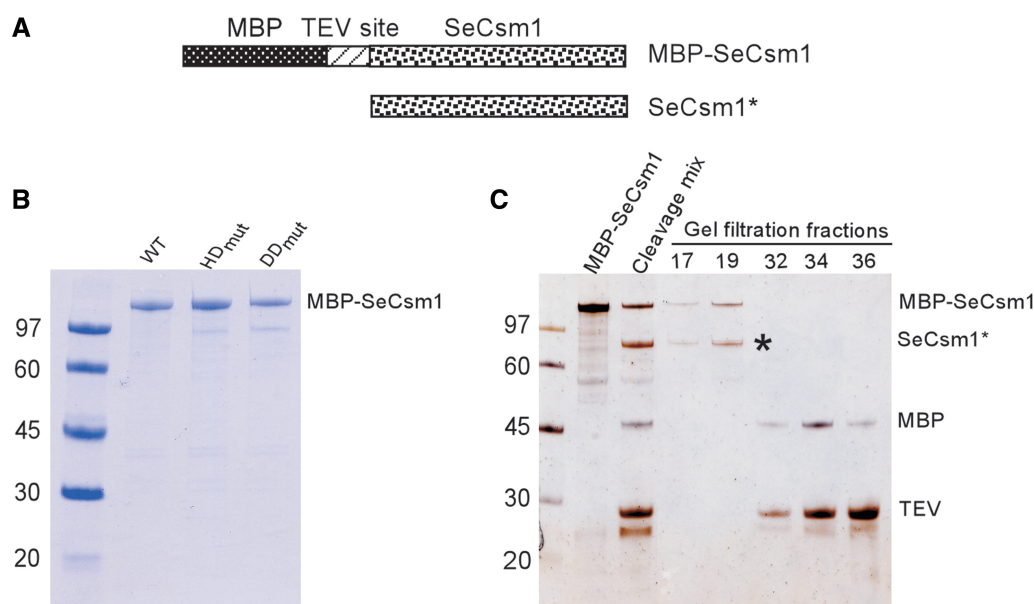


Figure 2. Purification of SeCsm1. (A) Schematic MBP-SeCsm1 chimera protein. (B) Sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis of purified wild-type MBP-SeCsm1 (WT), histidine–aspartate (HD) domain mutant (H14, D15 to asparagine, HD_{mut}) and GGDD mutant (D586, D587 to asparagine, DD_{mut}). (C) Silver-stained sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis of MBP removal from MBP-SeCsm1. The MBP tag was cleaved using TEV protease and the cleavage mixture was separated on a gel filtration column. Fraction numbers refer to those of the gel filtration elution. Fractions 17–19 were used for activity assays. The star denotes the tag-free SeCsm1 as confirmed by mass spectrometry analysis.

containing 20 mM Tris pH 7.5, 400 mM NaCl, 10% glycerol and 5 mM dithiothreitol (Figure 2B)

To remove the MBP fusion tag, the MBP-SeCsm1 sample was incubated with TEV protease overnight at 4°C. The reaction products were loaded on an analytical superose6 column (GE Healthcare) to separate the cleaved SeCsm1 from MBP, TEV and the uncleaved protein. Despite these efforts, homogenous SeCsm1 without MBP tag could not be completely separated from the uncleaved protein (Figure 2C). However, we confirmed the presence of the cleaved SeCsm1 protein by mass spectrometry analysis.

To create SeCsm1 mutants, the Pvp16-SERP2461 plasmid was mutated using the QuikChange kit

(Stratagene), and the mutant MBP-SeCsm1 and SeCsm1 proteins were purified in a way similar to that of the wild-type protein.

DNA and RNA cleavage assays

Single-stranded DNA and RNA substrates were purchased from Integrated DNA Technology (see Table 1 for sequence and length information). The DNA and RNA strands were 5'-radioactively-labeled with ³²P using the T4 polynucleotide kinase (Invitrogen) following the procedure of the manufacturer. The 3'-end-labeling of RNA substrates was performed using T4 RNA ligase (New England Biolabs) and cytidine-3',5'-bis-phosphate

Table 1. DNA and RNA oligos used for enzyme activity assays

Name (length)	Sequence
DNA1 (34 nts)	5'-TCGATGTAACGTATGCAAATGACAATTATTACTA-3'
DNA2 (34 nts)	5'-TAGTAATAAATTGTCATTTGCATACGTTACATCGA-3'
DNA3 (54 nts)	5'-CTACACTGTATCGATGTAACGTATGCAAATGACAATTATTACTACGAAGGCGAT-3'
DNA4 (54 nts)	5'-ATCGCCTTCG TAGTAATAAATTGTCATTTGCATACGTTACATCGA TACAGTGTAG-3'
DNA5 (40 nts)	5'-AATAATTGTCATTTGCATACGTTACATCGATACAGTGTAG-3'
DNA6 (40 nts)	5'-ATCGCCTTCGTAAGTAATAAATTGTCATTTGCATACGTTACA-3'
DNA7 (100 nts)	5'-TAATGAATAAGGACACGATTTGCCATCGCCTTCGTAGTAATAAATTGTCATTTGCATACGTTA CATCGATACAGTGTAGTAAAGCGTCATATGATTTAACG-3'
DNA8 (50 nts)	5'-ATGACAATTATTACTACGAAGGCGATGGCAATCGTGTCCCTTATTCATTA-3'
DNA9 (50 nts)	5'-CGTTAAATCATATGACGCTTTACTACACTGTATCGATGTAACGTATGCAA-3'
DNA10 (100 nts)	5'-CGTTAAATCATATGACGCTTTACTACACTGTATCGATGTAACGTATGCAAATGACAATTATTA CTACGAAGGCGATGGCAATCGTGTCCCTTATTCATTA-3'
DNA11 (110 nts)	5'-TAATGAATAAGGACACGATTTGCCATCGCCTTCGTTTTTACGTATGCCGAAGTATATAAATCA TCAGTACAAATTTTTTACAGTGTAGTAAAGCGTCATATGATTTAACG-3'
RNA1 (37 nts)	5'-ACGAGAACUAGUAAUUAUUUGUCAUUUGCAUACGUUAC-3'
RNA2 (43 nts)	5'-ACGAGAACUAGUAAUUAUUUGUCAUUUGCAUACGUUACAUUGAU-3'
RNA3 (30 nts)	5'-AUGCAAUAGACAAUUAUUACUAGUUCUCGU-3'
RNA4 (30 nts)	5'-AUCGAUGUAACGUAUGCAAUAGACAAUUAU-3'
RNA5 (43 nts)	5'-AUCGAUGUAACGUAUAGUGCAUUGUCCAGAUUACUAGUUCUCGU-3'

(Perkin Elmer). To generate double-stranded (ds) substrates used in the nuclease assays, the radioactively labeled DNA or RNA was incubated with the unlabeled complementary strand in a hybridization buffer containing 20 mM Tris pH 7.5 and 50 mM KCl. The mixture was incubated at 95°C for 15 min and then left to stand at room temperature. To form a bulge DNA substrate, DNA10 is annealed to DNA11, which creates a DNA duplex with 44 base mismatches in the middle, mimicking an R-loop formation.

To perform DNA or RNA cleavage, radioactively labeled DNA and RNA substrates were incubated for 1 h at 24°C in the presence or absence of the protein in a cleavage buffer containing 20 mM HEPES pH 7.0, 100 mM KCl, 2 mM MgCl₂ and (in the case of RNA substrates) 1 U of SUPERase-In Ribonuclease inhibitor (Applied Biosystems). The reaction was stopped by addition of the 2× formamide loading buffer. The mixture was then separated on a 15% denaturing polyacrylamide gel electrophoresis gel and the result was visualized using a phosphor screen (GE Healthcare).

RESULTS

Csm1 cleaves ssDNA and ssRNA

Csm1 nuclease activity was tested by adding purified MBP-SeCsm1 or a sample containing tag-free SeCsm1 protein to different substrates including ss- and dsDNA of different lengths, DNA–RNA hybrid, ssRNA and circular ssDNA (phage M13). No difference in cleavage activity was observed between MBP-SeCsm1 and SeCsm1. Thus, SeCsm1 is used for all subsequent discussions unless otherwise noted. In a cleavage buffer containing 20 mM HEPES (pH 7.0), 100 mM KCl, 2 mM MgCl₂ and 1 U of SUPERase-In Ribonuclease inhibitor, SeCsm1 has a clear cleavage activity on ssDNA (Figure 3A) and ssRNA (Figure 3B). SeCsm1 cleaved ssDNA and ssRNA irrespective of the sequence and length.

SeCsm1 requires divalent metals in cleaving nucleic acids

To identify divalent metals required for the cleavage activity of SeCsm1, we tested a panel of divalent metal ions that are known to be required by metal-dependent nucleases. Among the metals tested, only Mg²⁺ and Mn²⁺, but not Ni²⁺, Cu²⁺, Zn²⁺, Co²⁺, Ca²⁺ or Fe²⁺, supported the activity of SeCsm1. Furthermore, SeCsm1 appeared to be more active with Mg²⁺ than with Mn²⁺. As expected with metal-dependent nucleases, the cleavage reaction of SeCsm1 was completely inhibited by the metal-chelating agent ethylenediaminetetraacetic acid (Figure 3A and B).

SeCsm1 possesses a 3'-5' exonuclease activity

Based on the observed pattern of cleavage with SeCsm1, we hypothesized that SeCsm1 is an exonuclease. To test our hypothesis, we generated dsDNA substrates containing an overhang of the radioactively labeled DNA on either the 5'- or the 3'-end (Figure 4). Cleavage with SeCsm1 of two different DNAs showed that the substrate with its 3'-end protected was not cleaved, whereas that with its 5'-end protected generated a major product corresponding to the fragment without its unprotected 3' region (Figure 4A and B). The minor and shorter cleavage products observed for both DNA substrates (<10 nucleotides) likely were due to cleavage of unannealed DNA after continued incubation. This result shows that SeCsm1 is a DNA exonuclease in the 3'–5' direction.

Cleavage of RNA by SeCsm1 also revealed the 3'–5' exonuclease activity (Figure 4C). Short cleavage products were observed for the 5'- but not 3'-labeled substrate, although the 3'-labeled substrate was consumed by the enzyme in a time-dependent manner (Figure 4C). To further characterize the cleavage activity of SeCsm1 on RNA substrates, dsRNA substrates with 5' or 3' single-stranded overhangs, as well as bulged RNA duplexes labeled on either the 5'- or the 3'-end, were incubated

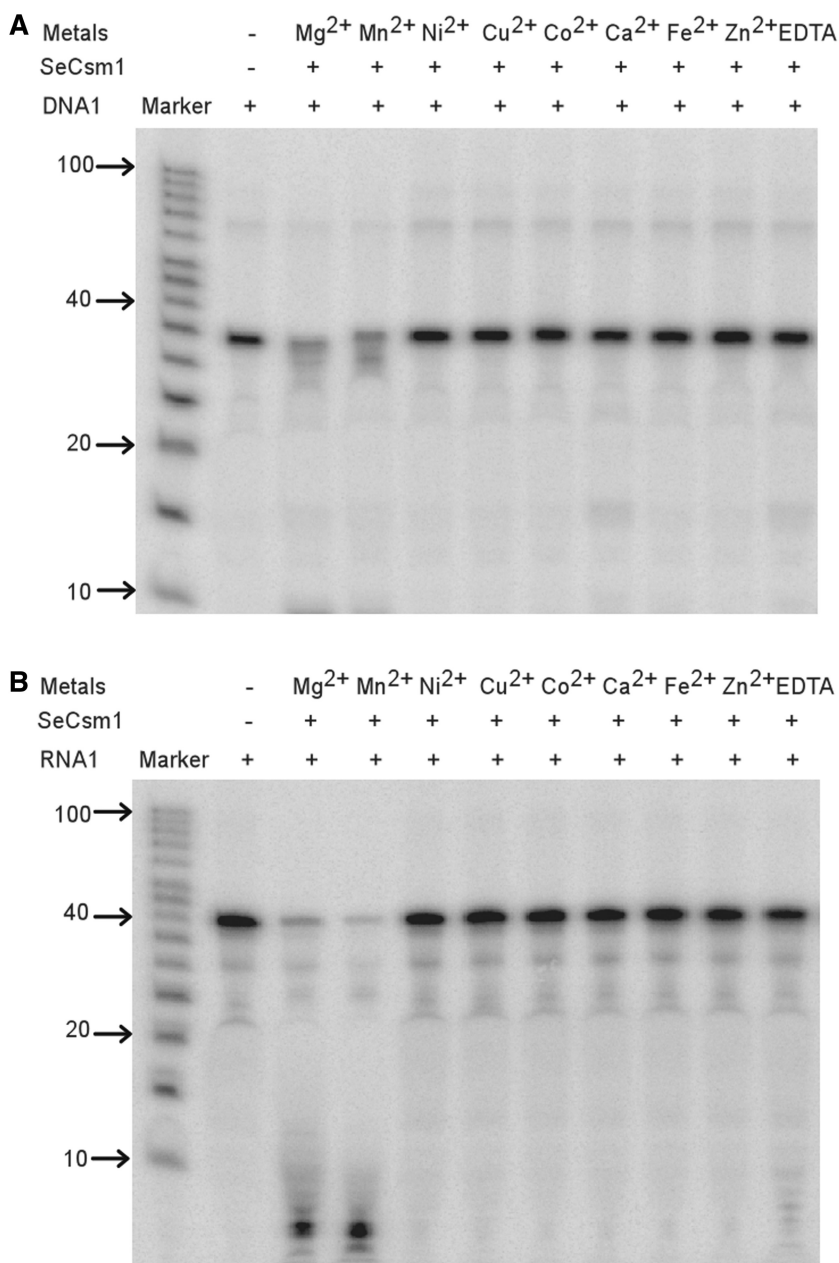


Figure 3. SeCsm1 cleaves single-stranded (ss)DNA and RNA. The sequences and length of the DNA and RNA substrates used are shown in Table 1. Lane 1 (labeled “Marker”) is an RNA marker in both panels. **(A)** The ssDNA cleavage activity of SeCsm1 and its metal dependence. SeCsm1 cleaves ssDNA in the presence of Mg²⁺ and Mn²⁺ but not other divalent cations tested. The metal-dependent cleavage activity is inhibited by the addition of ethylenediaminetetraacetic acid (EDTA). The concentration of the enzyme used in all reactions is 1 μ M, that of each divalent metal is 2 mM and that of EDTA is 10 mM. **(B)** The ssRNA cleavage activity of SeCsm1 and its metal dependence. Cleavage of ssRNA by SeCsm1 also depends on Mg²⁺ and Mn²⁺ and is inhibited by EDTA. Same concentrations are used as in (A).

with SeCsm1. The patterns of activity observed were more complex than that with DNA substrates and could not be simply explained by the 3′–5′ exonuclease activity. All three types of dsRNA showed cleavage products that were consistent with the 3′–5′ exonuclease and possibly an unknown endonuclease activity. The patterns, however, may also be explained by alternative secondary structures within the labeled strand (RNA2) that resulted in free 3′-end in all three cases. Whether SeCsm1 contains an endonuclease activity in addition to the 3′–5′

exonuclease activity requires a more extensive study with additional RNA substrates.

Consistent with the 3′–5′ exonuclease single-stranded activity on DNA, SeCsm1 showed no activity on a dsDNA or a bulged DNA structure generated by annealing two strands of 100 and 110 bases containing a central mismatched region of 44 bases (Figure 5A and B). Furthermore, SeCsm1 did not cleave the circular ssDNA of the M13 phage or a double-stranded plasmid DNA (Supplementary materials, Supplementary Figure S1).

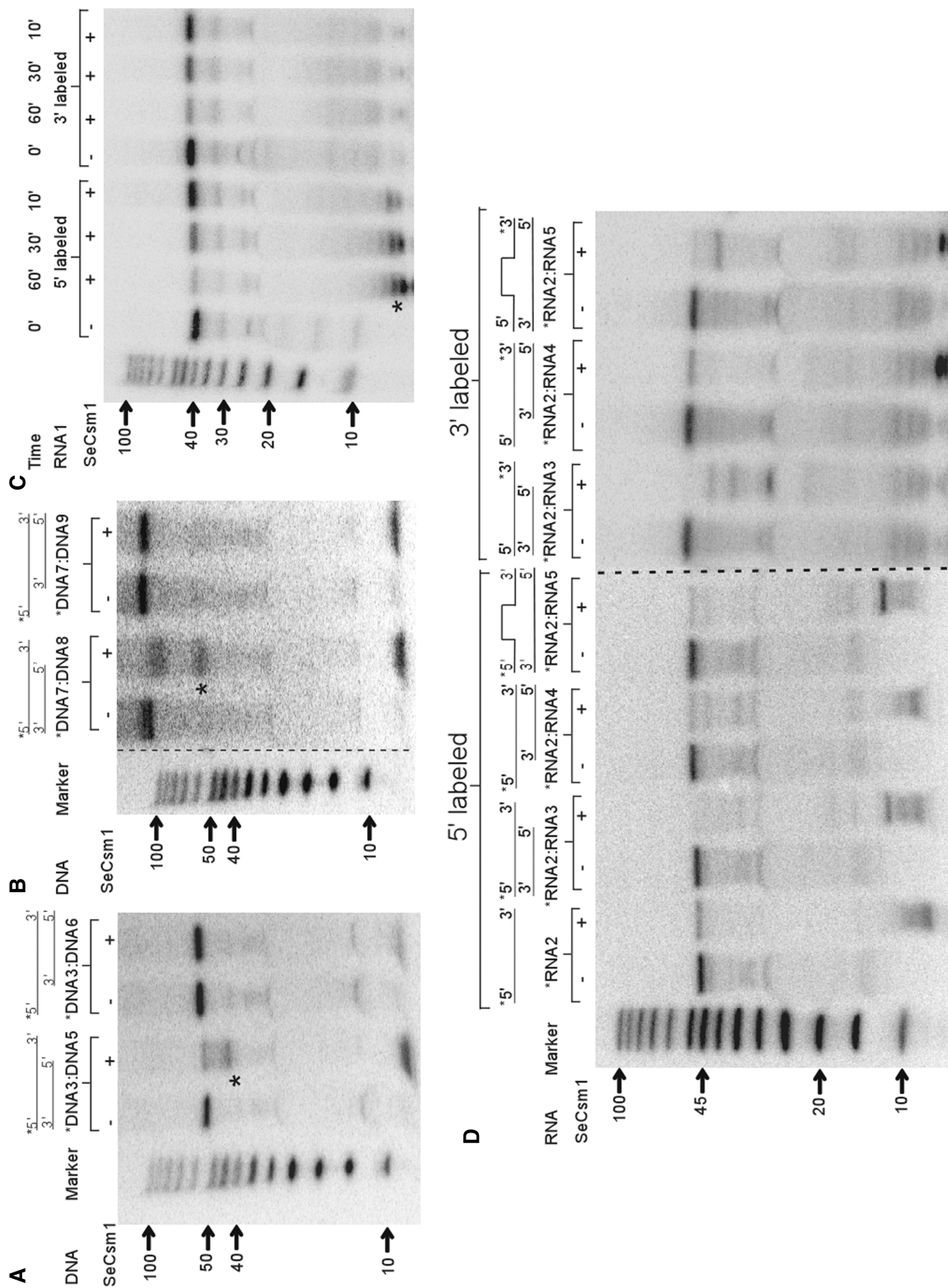


Figure 4. SeCsm1 is a 3'-5' exonuclease. DNA strands are annealed to form duplexes with either a 3' (DNA3:DNA5, DNA7:DNA8) or 5' overhang (DNA3:DNA6, DNA7:DNA9). The duplexes with a 3' overhang, but not 5' overhang, resulted in a major cleavage product consistent with the base-paired region (denoted with a star). Both types of duplexes exhibited shorter cleavage products that are believed to be due to cleavage of some unannealed ssDNA. The sequences and length of the DNA substrates used are shown in Table 1. RNA substrates have been labeled on either 5'- or 3'-ends, and tested for ssRNA cleavage in a time-dependent manner, or annealed to form duplexes (A) Cleavage of duplexes formed by DNA3, DNA5 and DNA6 (Table 1). (B) Cleavage of duplexes formed by DNA7, DNA8 and DNA9. (C) Time-course study of the cleavage of SeCsm1 on RNA1 labeled on the 5'- or 3'-end. (D) Cleavage results of SeCsm1 on RNA strands with either 3' or 5' overhangs and a bulged structure. The RNA2 strand has been labeled on either the 3'- or 5'-end. RNA oligos used to form these substrates are described in Table 1. The first lane is a 5'-labeled RNA marker.

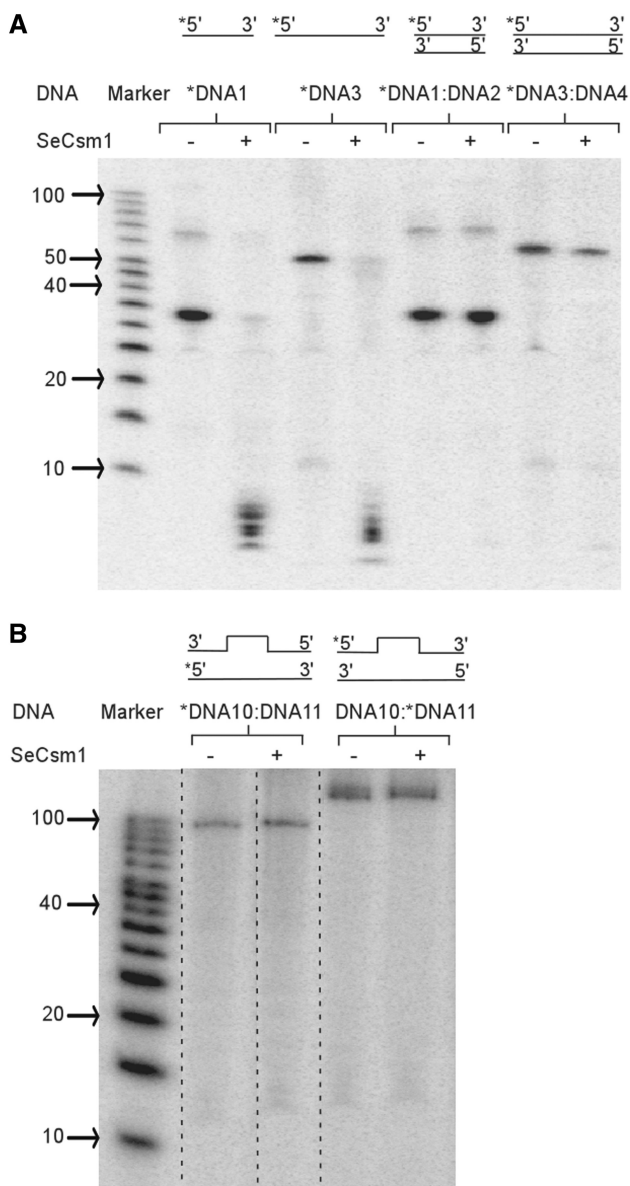


Figure 5. SeCsm1 does not have a DNA endonuclease cleavage activity. Sequences of the DNA substrates used are shown in Table 1. The first lane is a 5'-labeled RNA marker. (A) SeCsm1 cleaves single-stranded DNA1 and DNA3 (lanes 3 and 5) but not the DNA1:DNA3 duplex (lanes 7 and 9). (B). SeCsm1 does not cleave the bulged duplex DNA mimicking an artificial R-loop. The DNA bulge is formed by annealing DNA 10 and DNA11, where either DNA10 or DNA11 is radiolabeled.

The GGDD motif of the cyclase-like domain contains the active site

To locate the active site of SeCsm1 responsible for the DNA cleavage activity, we performed site-directed mutagenesis of residues potentially involved in catalysis within both the HD and the cyclase-like domains (Figure 6A). We mutated the signature residues, His14 and Asp15, of the HD domain to asparagine to create HD_{mut}. These two residues likely participate in the coordination of metal ions as found in HD proteins (34,35). We also mutated

the residues, Asp586 and Asp587, of the cyclase-like domain to asparagine to create DD_{mut}. The corresponding aspartates are part of the GGDD motif in the homologous protein Cmr2 that binds two Ca²⁺ ions (43) (full-sequence alignment of SeCsm1 and Cmr2 present in Supplementary Figure S2). A similar motif in nucleotidyl cyclases containing GGDEF forms the catalytic center for the nucleotide cyclization reaction (46). The GGDD motif in Cmr2 is thus a degenerate GGDEF motif and has no known cyclization activity (43). All four mutated residues of SeCsm1, His14, Asp15, Asp586 and Asp587 are highly conserved in more than 100 Csm1 sequences examined. Based on the similar gel filtration elution patterns, we believe that these mutations did not disrupt folding of SeCsm1 (data not shown).

DNA cleavage assay with the wild-type and the mutants showed that while HD_{mut} had a reduced activity, DD_{mut} completely abolished ssDNA cleavage activity (Figure 6B). When tested on ssRNA substrate, HD_{mut} generated a similar cleavage pattern as the wild-type SeCsm1, while DD_{mut} resulted in less number of products (Figure 6C), suggesting that the GGDD motif is also an RNA cleavage center. The common product observed in both HD_{mut} and DD_{mut} cleavage is likely due to an alternative cleavage site. These results suggest that although the GGDD motif is directly responsible for the 3'-5' exonuclease activity of SeCsm1 on ssDNA and ssRNA, the HD domain likely takes part in this process.

DISCUSSION

Staphylococcus epidermidis strain RP62a is one of the first organisms where interference with horizontal gene transfer was described and its CRISPR system was found to target DNA (10). Yet the molecular mechanism of how this organism destroys invading DNA remains uncharacterized. SeCsm1 is a hallmark protein of this organism that is believed to be involved in DNA interference (16,17) and crRNA processing (21). SeCsm1 is classified as a member of the Cas10 family of proteins that are characterized by a permuted HD domain and a nucleotidyl cyclase-like domain adopting a distinct RNA Recognition Motif fold (47). Although both domains likely have divalent metal-binding sites, their roles in nucleic acid cleavage remain undefined.

We showed that SeCsm1 cleaves ssDNA and ssRNA. This result is in agreement with the previously predicted nuclease activity in SeCsm1 (7,17,18), and with the finding that SeCsm1 may be involved in the maturation of crRNA (21). More recently, Csm1 is found to form a complex termed Cas10/Csm with Csm2-5 and the crRNA comparable with Cascade and the Type III-B Cmr complexes (48). This complex is thought to mediate immunity against invading DNA and crRNA maturation in *S. epidermidis*. The possible nuclease(s), however, involved in silencing DNA and processing crRNA remain unknown. The CRISPR system of *S. epidermidis* is a 'minimal CRISPR system' and the nuclease activity of Csm1 could be used for generation of mature crRNA as well as silencing the invading DNA. The 3'-5' DNA exonuclease activity of

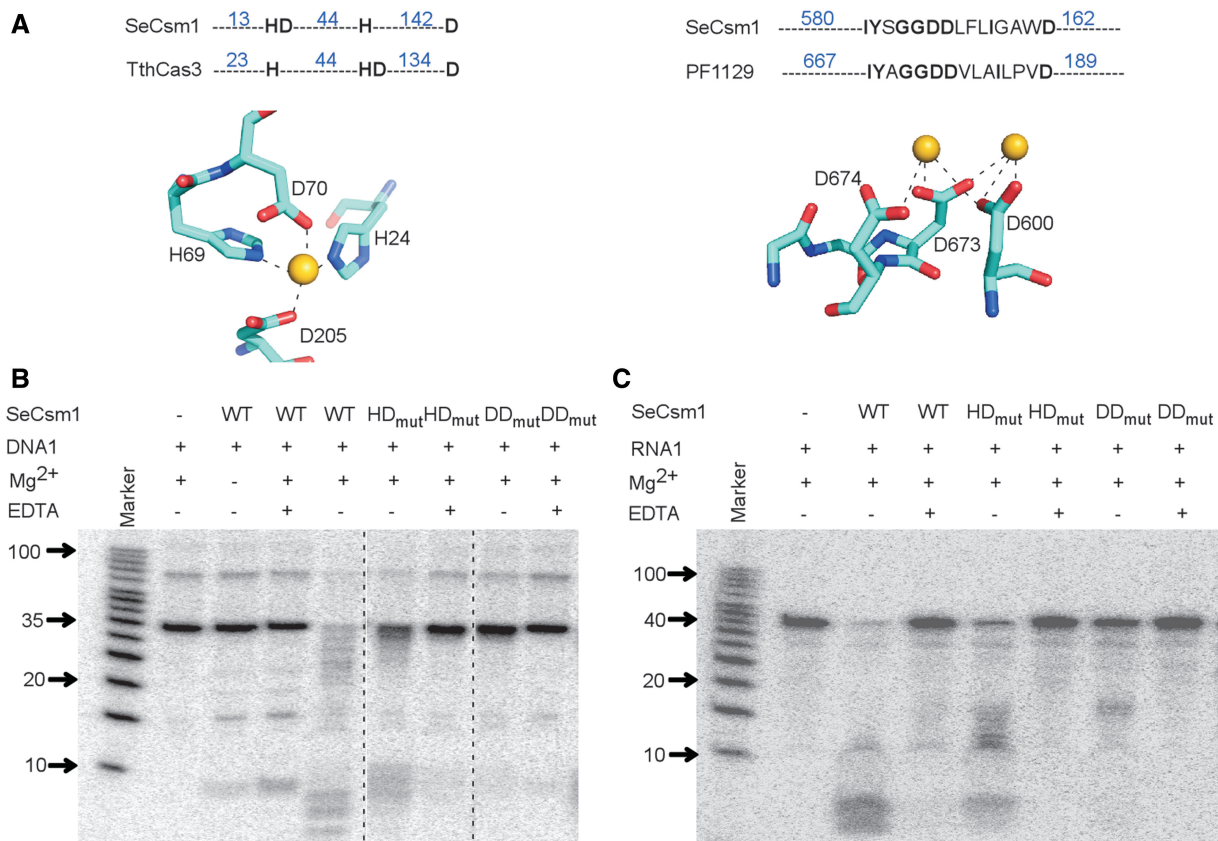


Figure 6. The GGDD motif of the cyclase-like domain of SeCsm1 is involved in the nuclease activity. (A) Sequence features and representative structure of a known HD domain from *Thermus thermophilus* Cas3 (TTHB187) (left) and those of the cyclase-like domain from *Pyrococcus furiosus* Cmr2 (Pfl129) (right). The bound metal ions in both structures are represented by orange spheres and residues coordinated with the metals are labeled. Note that the HD domain of SeCsm1 has a permuted catalytic residue arrangement as compared with that of Cas3, whereas its GGDD motif shares the same sequence pattern as the GGDD motif of Cmr2. The numbers in the sequence alignments represent the numbers of residues not shown. (B) Comparison of wild-type SeCsm1 (WT) and its mutants (DD_{mut} and HD_{mut}) in ssDNA cleavage activity. (C) Comparison of the wild-type SeCsm1 (WT) and its mutants (DD_{mut} and HD_{mut}) in ssRNA cleavage activity.

SeCsm1 may be used in the context of the intact effector complex to cleave the invading DNA after it has been nicked by another endonuclease.

The discovery that the GGDD motif within the cyclase-like domain rather than the signature HD domain of SeCsm1 exhibits a nuclease activity is surprising to us. First, the GGDD motif of the Type III-B Cas10 protein, Cmr2, was found to have no endoribonuclease activity (43). Second, multiple HD proteins or HD domains have been found to possess ssDNA cleavage activity (33–35). The fact that SeCsm1 GGDD is mainly responsible for the observed DNA cleavage suggests that this motif may have an intrinsic nucleic acid cleavage activity and that this activity may depend on the context of the entire protein. Indeed, despite both being members of the Cas10 family, Csm1 and Cmr2 share a low sequence homology, suggesting differences in other regions of the proteins. To try to explain why the HD domain of SeCsm1 is not critical to its nuclease activity, we note that the HD domain of Csm1 has a permuted arrangement of HD...H...D (Figure 6A), whereas a typical HD domain has the arrangement of H...HD...D (49).

The nuclease activity of SeCsm1 observed is metal-ion-dependent and the enzyme is active in the presence of Mg²⁺ and Mn²⁺ but not Cu²⁺, Zn²⁺ or Ni²⁺. The similar dependence on divalent metals has been observed in other CRISPR nucleases that include the HD proteins (33–35), Cas1 (50), Cas2 (51,52) and Cas5d (31). We note that none of these previously known CRISPR nuclease proteins include the GGDD motif. Thus our discovery adds another important protein motif capable of cleaving nucleic acids.

Although our results show that SeCsm1 cleaves DNA and RNA non-specifically, it is plausible that the required specificity for DNA and RNA is conferred by other Csm proteins and the bound crRNA of the Cas10/Csm complex. During DNA silencing, one or more Csm subunits (Csm2–Csm5) may be required to produce the first cut in the DNA, and the exonuclease activity of Csm1 further digests the DNA with free 3'-end. Alternatively, the Csm complex might have a surveillance function similar to Cascade, and when it encounters an invading DNA complementary to the crRNA, another protein is recruited to perform the actual degradation process. Further studies of the Csm proteins are

required to distinguish these models and to uncover the mechanism of DNA silencing in *S. epidermidis*.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Funding for open access charge: National Institutes of Health R01 [GM099604 to H.L.].

Conflict of interest statement. None declared.

REFERENCES

- Richter,C., Chang,J.T. and Fineran,P.C. (2012) Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated (Cas) systems. *Viruses*, **4**, 2291–2311.
- Hendrix,R.W. (2003) Bacteriophage genomics. *Curr. Opin. Microbiol.*, **6**, 506–511.
- Wommack,K.E. and Colwell,R.R. (2000) Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.*, **64**, 69–114.
- Weinbauer,M.G. (2004) Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.*, **28**, 127–181.
- Fineran,P.C. and Charpentier,E. (2012) Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology*, **434**, 202–209.
- Labrie,S.J., Samson,J.E. and Moineau,S. (2010) Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.*, **8**, 317–327.
- Makarova,K.S., Grishin,N.V., Shabalina,S.A., Wolf,Y.I. and Koonin,E.V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct.*, **1**, 7.
- Barrangou,R., Fremaux,C., Deveau,H., Richards,M., Boyaval,P., Moineau,S., Romero,D.A.P. and Horvath,P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709–1712.
- 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 der Oost,J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.
- Marraffini,L.A. and Sontheimer,E.J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, **322**, 1843–1845.
- Hale,C.R., Zhao,P., Olson,S., Duff,M.O., Graveley,B.R., Wells,L., Terns,R.M. and Terns,M.P. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, **139**, 945–956.
- Bolotin,A., Quinquis,B., Sorokin,A. and Ehrlich,S.D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, **151**(Pt 8), 2551–2561.
- Mojica,F.J., Diez-Villaseñor,C., García-Martínez,J. and Soria,E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.*, **60**, 174–182.
- Jansen,R., Embden,J.D., Gastra,W. and Schouls,L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.*, **43**, 1565–1575.
- Pourcel,C., Salvignol,G. and Vergnaud,G. (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, **151**(Pt 3), 653–663.
- Haft,D.H., Selengut,J., Mongodin,E.F. and Nelson,K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.*, **1**, e60.
- Makarova,K.S., Aravind,L., Wolf,Y.I. and Koonin,E.V. (2011) Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Direct.*, **6**, 38.
- Makarova,K.S., Haft,D.H., Barrangou,R., Brouns,S.J., Charpentier,E., Horvath,P., Moineau,S., Mojica,F.J., Wolf,Y.I., Yakunin,A.F. et al. (2011) Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.*, **9**, 467–477.
- Carte,J., Wang,R., Li,H., Terns,R.M. and Terns,M.P. (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.*, **22**, 3489–3496.
- Haurwitz,R.E., Jinek,M., Wiedenheft,B., Zhou,K. and Doudna,J.A. (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science*, **329**, 1355–1358.
- Hatoum-Aslan,A., Maniv,I. and Marraffini,L.A. (2011) Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc. Natl Acad. Sci. USA*, **108**, 21218–21222.
- Przybilski,R., Richter,C., Gristwood,T., Clulow,J.S., Vercoe,R.B. and Fineran,P.C. (2011) Csy4 is responsible for CRISPR RNA processing in *Pectobacterium atrosepticum*. *RNA Biol.*, **8**, 517–528.
- Wiedenheft,B., Lander,G.C., Zhou,K., Jore,M.M., Brouns,S.J., van der Oost,J., Doudna,J.A. and Nogales,E. (2011) Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature*, **477**, 486–489.
- Westra,E.R., van Erp,P.B., Künne,T., Wong,S.P., Staals,R.H., Seegers,C.L., Bollen,S., Jore,M.M., Semenova,E., Severinov,K. et al. (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell.*, **46**, 595–605.
- Zhang,J., Rouillon,C., Kerou,M., Reeks,J., Brugger,K., Graham,S., Reimann,J., Cannone,G., Liu,H., Albers,S.V. et al. (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol. Cell*, **45**, 303–313.
- Deltcheva,E., Chylinski,K., Sharma,C.M., Gonzales,K., Chao,Y., Pirzada,Z.A., Eckert,M.R., Vogel,J. and Charpentier,E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, **471**, 602–607.
- Gasiunas,G., Barrangou,R., Horvath,P. and Siksnys,V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA*, **109**, E2579–E2586.
- Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
- Sinkunas,T., Gasiunas,G., Waghmare,S.P., Dickman,M.J., Barrangou,R., Horvath,P. and Siksnys,V. (2013) *In vitro* reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J.*, **32**, 385–394.
- 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.
- Nam,K.H., Haitjema,C., Liu,X., Ding,F., Wang,H., DeLisa,M.P. and Ke,A. (2012) Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulgr CRISPR-Cas system. *Structure*, **20**, 1574–1584.
- van der Oost,J., Jore,M.M., Westra,E.R., Lundgren,M. and Brouns,S.J. (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.*, **34**, 401–407.
- Sinkunas,T., Gasiunas,G., Fremaux,C., Barrangou,R., Horvath,P. and Siksnys,V. (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.*, **30**, 1335–1342.
- Mulepati,S. and Bailey,S. (2011) Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). *J. Biol. Chem.*, **286**, 31896–31903.

35. Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A. and Yakunin, A.F. (2011) Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO J.*, **30**, 4616–4627.
36. Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R. *et al.* (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.*, **18**, 529–536.
37. Westra, E.R., Nilges, B., van Erp, P.B., van der Oost, J., Dame, R.T. and Brouns, S.J. (2012) Cascade-mediated binding and bending of negatively supercoiled DNA. *RNA Biol.*, **9**, 1134–1138.
38. Howard, J.A., Delmas, S., Ivančić-Baće, I. and Bolt, E.L. (2011) Helicase dissociation and annealing of RNA-DNA hybrids by *Escherichia coli* Cas3 protein. *Biochem. J.*, **439**, 85–95.
39. Karvelis, T., Gasiunas, G., Miksys, A., Barrangou, R., Horvath, P. and Siksnys, V. (2013) crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA Biol.*, **10**, 841–851.
40. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
41. Chang, N., Sun, C., Gao, L., Zhu, D., Xu, X., Zhu, X., Xiong, J.W. and Xi, J.J. (2013) Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res.*, **23**, 465–472.
42. Anantharaman, V., Iyer, L.M. and Aravind, L. (2010) Presence of a classical RRM-fold palm domain in Thg1-type 3'-5' nucleic acid polymerases and the origin of the GGDEF and CRISPR polymerase domains. *Biol. Direct.*, **5**, 43.
43. Cocozaki, A.I., Ramia, N.F., Shao, Y., Hale, C.R., Terns, R.M., Terns, M.P. and Li, H. (2012) Structure of the Cmr2 subunit of the CRISPR-Cas RNA silencing complex. *Structure*, **20**, 545–553.
44. Shao, Y., Cocozaki, A.I., Ramia, N.F., Terns, R.M., Terns, M.P. and Li, H. (2013) Structure of the Cmr2-Cmr3 subcomplex of the Cmr RNA silencing complex. *Structure*, **21**, 376–384.
45. Marraffini, L.A. and Sontheimer, E.J. (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, **463**, 568–571.
46. Pei, J. and Grishin, N.V. (2001) GGDEF domain is homologous to adenylyl cyclase. *Proteins*, **42**, 210–216.
47. Makarova, K.S., Aravind, L., Grishin, N.V., Rogozin, I.B. and Koonin, E.V. (2002) A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res.*, **30**, 482–496.
48. Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W. and Marraffini, L.A. (2013) A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *J. Biol. Chem.*, **288**, 27888–27897.
49. Aravind, L. and Koonin, E.V. (1998) The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem. Sci.*, **23**, 469–472.
50. Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S.M., Ma, W. and Doudna, J.A. (2009) Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure*, **17**, 904–912.
51. Beloglazova, N., Brown, G., Zimmerman, M.D., Proudfoot, M., Makarova, K.S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W. *et al.* (2008) A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol. Chem.*, **283**, 20361–20371.
52. Nam, K.H., Ding, F., Haitjema, C., Huang, Q., DeLisa, M.P. and Ke, A. (2012) Double-stranded endonuclease activity in *Bacillus halodurans* clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. *J. Biol. Chem.*, **287**, 35943–35952.