Allosteric regulation of Csx1, a type IIIB-associated CARF domain ribonuclease by RNAs carrying a tetraadenylate tail
Han, Wenyuan; Pan, Saifu; López-Méndez, Blanca; Montoya, Guillermo; She, Qunxin

Published in: Nucleic Acids Research

DOI: 10.1093/nar/gkx726

Publication date: 2017

Document version Publisher's PDF, also known as Version of record

Document license: CC BY-NC

Citation for published version (APA): Han, W., Pan, S., López-Méndez, B., Montoya, G., & She, Q. (2017). Allosteric regulation of Csx1, a type IIIB-associated CARF domain ribonuclease by RNAs carrying a tetraadenylate tail. Nucleic Acids Research, 45(18), 10740-10750. [gkx726]. https://doi.org/10.1093/nar/gkx726
Allosteric regulation of Csx1, a type IIIB-associated CARF domain ribonuclease by RNAs carrying a tetraadenylate tail

Wenyuan Han1, Saifu Pan2, Blanca López-Méndez3, Guillermo Montoya4 and Qunxin She1,2,*

1Archea Center, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, Copenhagen Biocenter, DK-2200 Copenhagen N, Denmark, 2State Key Laboratory of Agricultural Microbiology and College of Life Science and Technology, Huazhong Agricultural University, 430070 Wuhan, China, 3Protein Structure & Function Programme, Protein Production and Characterization Platform, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen, Denmark and 4Macromolecular Crystallography Group, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen, Denmark

ABSTRACT

CRISPR–Cas systems protect prokaryotes against invading viruses and plasmids. The system is associated with a large number of Cas accessory proteins among which many contain a CARF (CRISPR-associated Rossmann fold) domain implicated in ligand binding and a HEPN (higher eukaryotes and prokaryotes nucleotide-binding) nuclease domain. Here, such a dual domain protein, i.e. the Sulfolobus islandicus Csx1 (SisCsx1) was characterized. The enzyme exhibited metal-independent single-strand specific ribonuclease activity. In fact, SisCsx1 showed a basal RNase activity in the absence of ligand; upon the binding of an RNA ligand carrying four continuous adenosines at the 3′-end (3′-tetra-rA), the activated SisCsx1 degraded RNA substrate with a much higher turnover rate. Amino acid substitution mutants of SisCsx1 were obtained, and characterization of these mutant proteins showed that the CARF domain of the enzyme is responsible for binding to 3′-tetra-rA and the ligand binding strongly activates RNA cleavage by the HEPN domain. Since RNA polyadenylation is an important step in RNA decay in prokaryotes, and poly(A) RNAs can activate CARF domain proteins, the poly(A) RNA may function as an important signal in the cellular responses to viral infection and environmental stimuli, leading to degradation of both viral and host transcripts and eventually to cell dormancy or cell death.

INTRODUCTION

Clustered regularly interspaced short palindromic repeats (CRISPR) loci and CRISPR-associated (Cas) proteins is the inheritable and adaptive immune system that protects bacteria and archaea from invasion by viruses and plasmids (1,2). The system is able to acquire a short fragment from foreign genetic elements and insert into the CRISPR loci as a new spacer. Then transcription of CRISPR loci and processing of the transcripts generate CRISPR RNA (crRNA) based on each spacer. Finally, crRNA guides Cas proteins (a protein complex or a single protein) to recognize and destroy invading DNA and/or RNA. Therefore, there are three basic steps in the CRISPR immunity: new spacer acquisition, biogenesis of crRNA and interference of invading nucleic acid (1–5).

In addition, CRISPR–Cas systems are usually associated with genes that appear to be not directly involved in the three steps (5). The proteins encoded by such genes, usually termed non-core Cas accessory proteins, are still poorly characterized. The most widespread group of accessory proteins contain a predicted nucleotide-binding domain with a Rossmann-like fold and are denoted CARF (CRISPR-associated Rossmann fold) domain proteins (6). In the proteins belonging to this group, the CARF domain is usually fused to another effector domain, such as a winged helix-turn-helix (WHH) DNA binding domain, a DNase domain of the restriction endonuclease or an RNase domain belonging to either RelE or HEPN (higher eukaryotes and prokaryotes nucleotide-binding) families (6-9). Since Rossmann folds are known to function in ligand binding, it is predicted that the CARF domain could have an allosteric regulatory function upon binding to a ligand molecule. In the crystal structures solved for a few CARF domain proteins, a ligand-binding site is predicted in

*To whom correspondence should be addressed. Tel: +45 3532 2013; Fax: +45 3532 2128; Email: qunxin@bio.ku.dk

© The Author(s) 2017. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
the CARF domain containing several amino acids residues (10–12), and furthermore, a co-crystal structure has been solved for the Archaeoglobus fulgidus Csx3, a single domain CARF protein and a tetranucleotide that binds to the protein (12).

Within the CARF group, Csm6 and Csx1 proteins are encoded by the genes evolutionarily linked with type III CRISPR–Cas systems (4,13), the only type of CRISPR–Cas system that performs dual RNA/DNA targeting activity (14–24). Nevertheless, Csx1 and Csm6 are not found to be associated with any type III effector complexes. Furthermore, they are not required for either RNA cleavage or DNA cleavage by type III effector complexes in vitro (14,16,18,20–22). Although it has been suggested that Csx1/Csm6 is necessary for the transcription-dependent plasmid silencing by Cmr-α in Sulfolobus islandicus and Csm in Staphylococcus epidermidis respectively (15,25), new evidence shows that they are dispensable for the in vivo DNA targeting by several type III systems (21,26).

Structural and biochemical analyses of Csm6 and Csx1 proteins show that they contain a C-terminal HEPN domain and representative proteins, including Pyrococcus furiosus Csx1 and Thermus thermophilus Csm6, exhibit metal-independent single-strand (ss)-specific endoribonuclease activity (11,27). It is believed that the ribonuclease activity is performed by the HEPN domain, while the functions of the CARF domain are not understood yet (11,27).

To reveal whether and how CARF domain of Csx1 performs ligand binding and allosteric regulatory function, we characterized the recombinant Csx1 encoded in S. islandicus Rey15A (SisCsx1). The results show that the ssRNA cleavage activity of SisCsx1 is activated by 3′-teta-rA (3′-end four continuous adenosine nucleotides). We further demonstrate that the CARF domain binds to 3′-teta-rA and regulates the RNase activity of HEPN domain.

**MATERIALS AND METHODS**

**Strains and growth conditions**

Genetic hosts *S. islandicus* E233S1 and the *cxl* deletion mutant (Δcsx1) were grown in SCVU medium (basic salts and 0.2% sucrose, 0.2% cas amino acids, 1% vitamin solution, 0.2% uracil) at 78°C, while uracil was removed from the medium for transformants. Transformation and construction of Δcsx1 was performed as described previously (28). The primers used for construction of Δcsx1 are listed in Supplementary Table S1. *Escherichia coli* DH5α and Rosetta strain were used for plasmid construction and recombinant protein expression in *E. coli*, respectively.

**Construction of expression plasmids**

To construct pSeSD-based expression plasmid, wild type (WT) *cxl* gene was amplified with the primers SiRe_0884-up-NheI and SiRe_0884-dw-Sall (Supplementary Table S1) and inserted into pSeSD between NheI and Sall. To construct pET30a-based plasmid, the pSeSD-csx1 plasmid was digested by NdeI and Sall, and the gene fragment was purified and inserted into pET30a vector between NdeI and XhoI. Overlapping-PCR was employed to produce *cxl* mutants with the primers listed in Supplementary Table S1, except Csx1M3, which was amplified with SiRe_0884M3-up-NheI and SiRe_0884-dw-Sall (Supplementary Table S1). Then, the mutated gene fragments were inserted into pSeSD and pET30a as the procedure for the wild type gene.

All the primers to be used for DNA cloning were synthesized from Integrated DNA Technology (IDT, USA). Sequences of all plasmid constructs were verified by DNA sequencing at MacroGen Europe (Amsterdam, Netherlands).

**Expression and purification of Csx1 and the mutants**

The pSeSD-based expression plasmid was transformed into the genetic host E233S1 of *S. islandicus* Rey15A to express WT Csx1, or to Δcsx1 to express the Csx1 mutants, respectively. The transformants were grown in SCV medium until the optical density of ~0.2 and then the expression was induced by 0.2% arabinose. However, the expression of Csx1M2 from *S. islandicus* Rey15A was not successful. Therefore, we expressed WT, Csx1M2, M3, M4, M5 and M6 in *E. coli* for this purpose, the pET30a-based expression plasmids were transformed into *E. coli* Rosetta strain, and the expression was induced by 2.5% lactose at room temperature for overnight.

The purification procedure reported for the purification of native Cmr-α complex was followed with modification (24). Specifically, cell pellet, either *S. islandicus* cells or *E. coli* cells, was re-suspended in Buffer A (20 mM HEPES pH 7.5, 30 mM Imidazole, 500 mM NaCl) and disrupted by French press. The cell extract was loaded onto a 1 ml His-Trap HP (GE) and His-tagged protein was eluted by Buffer B (20 mM HEPES pH 7.5, 500 mM Imidazole, 500 mM NaCl). The Buffer B fractions were concentrated and further purified by size exclusion chromatography in Buffer C (20 mM Tris–HCl pH 7.5, 500 mM NaCl) with a Superdex 200 10/300 GL column (GE). The fractions containing the target protein were pooled, concentrated and used for further analysis.

**Size-exclusion chromatography–multi-angle light scattering**

The oligomeric state of Csx1 was analyzed by size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS). Protein samples were prepared at 1 mg/ml concentration and dialyzed into gel filtration running buffer, 20 mM Tris–HCl, 150 mM NaCl, 0.5 mM TCEP, pH 8.0. The samples were loaded on a Superdex 200 Increase 10/300 size exclusion column (GE Healthcare). The column outlet was directly connected to a DAWN HELEOS II MALS detector (Wyatt Technology) followed by an Optilab T-rEX differential refractometer (Wyatt Technology). Data were collected and analysed using ASTRA 6 software (Wyatt Technology). Samples were run in triplicates. The monomeric BSA was used as standard (Sigma-Aldrich).

**Labeling of DNA and RNA substrates**

RNA and DNA substrates used in cleavage assays were 5′ labeled with 32P using T4 polynucleotide kinase (New England Biolabs). Double strand (ds) DNA was generated by annealing of labeled SS1 ssDNA with unlabeled SS1T ssDNA, while dsRNA was made by annealing of labelled
SS1–40 and SS1–40T (Supplementary Table S2). Then, all substrates were purified by cutting gel from native (dsDNA and dsRNA) or denaturing (ssDNA/RNA) PAGE. The RNA oligos used in the UV crosslink assay were also labeled in a similar way and diluted to 500 nM. Further, the RNA oligos used in the UV crosslink assay have been analyzed by denaturing PAGE to make sure the radioactivity of them was identical. All the DNA and RNA oligos were purchased from IDT, USA.

**RNA cleavage assay**

To analyze the metal-dependency of SisCsx1-mediated RNA cleavage, labeled SS1–40 was incubated with 3 μM SisCsx1 in the buffer containing 20 mM Mes (pH 6.2) and 5 mM DTT, as well as indicated metal ions or EDTA. For RNA cleavage experiments, the reactions were carried out in the buffer containing 20 mM Mes (pH 6.2), 5 mM MgCl₂, 5 mM MnCl₂ and 5 mM DTT, as well as indicated substrates, enzyme and unlabeled nucleic acids. The RNA cleavage activity of Cmr-α was performed as previously described (24). To analyze the RNA degradation of a mixture of Cmr-α and Csx1, the two proteins were pre-mixed and then added into the reaction mixture. All the reactions were performed at 70°C and stopped at the indicated time point by supplementing 2 × RNA loading dye (New England Biolabs) and cooling on ice. At last, the samples were heated for 3 min at 95°C and then separated by 18% polyacrylamide denaturing gel and visualized by phosphor imaging. RNA ladders were generated by Decade™ Marker RNA (ambion) following the instructions.

**UV crosslink assay**

For the UV crosslinking assay, labeled different RNA oligos were incubated with 3 μM SisCsx1 or mutants in the buffer containing 20 mM Mes (pH 6.2), 5 mM MgCl₂, 5 mM MnCl₂ and 5 mM DTT at 70°C for 10 min, and then exposed to UV irradiation (254 nm) for 30 min in UV stratalinker 1800 (stratagene, USA). The samples were resolved by SDS-loading buffer and heated at 70°C for 5 min. At last, the samples were loaded onto 12% SDS-polyacrylamide gel, and the radioactive signal was visualized by phosphor imaging while the protein band was visualized by Coomassie staining.

**Poly(A) polymerase extension assay**

To analyze the 3′-end group of the cleavage products generated by SisCsx1, 5′-labeled SS1–40 substrate was incubated with SisCsx1 in the presence of S3A7 or not at 70°C for 20 min. Then, half of the reaction mixture was treated with 5 U poly(A) polymerase (PAP, New England Biolabs), 1 mM ATP and 1 × PAP reaction buffer at 37°C for 10 min, while the rest of the reaction mixture was incubated with 1 mM ATP and 1 × PAP reaction buffer as controls. At last, the reactions were stopped by addition of 2 × RNA loading dye (New England Biolabs), heated at 95°C for 3 min and analyzed by 18% polyacrylamide denaturing gel.

**RESULTS**

The *S. islandicus* Csx1 is a metal-independent single-strand endoribonuclease

The *S. islandicus* csx1 gene is located adjacent to the *cas* gene cassette of the III-B Cmr-α system and encodes a Cas accessory protein of 454 amino acids (Supplementary Figure S1). To characterize the encoded protein, the gene was cloned into the *Sulfolobus* expression vector pSesSD, from which a recombinant SisCsx1 protein carrying a 6xHis peptide at the N terminus was expressed. SDS-PAGE analysis of purified SisCsx1 revealed a single protein band of ca. 50 kDa (Figure 1A), while Multi-Angle Laser Light Scattering (MALLS) analysis indicates that SisCsx1 has a size of ~255 kDa in solution (Figure 1B). The molecular weight of our construct is 53 kDa, suggesting that SisCsx1 could be in a pentameric configuration; however, further analysis is needed to determine the stoichiometry of the oligomer.

RNA cleavage activity of SisCsx1 was tested with the SS1–40 substrate, a single-strand (ss) RNA (RNA protospe1 of *lacS* gene) that has been used for characterization of the Cmr-α effector complex (24). To ensure the detection of the RNase activity, a relatively large amount of the purified SisCsx1 protein (3 μM) was mixed with labeled SS1–40 in the presence of Mg²⁺ or Mn²⁺, or EDTA. The reactions were incubated at 70°C for 60 min, and RNA cleavage products were analyzed by denaturing gel electrophoresis. As shown in Figure 1C, SisCsx1 cleaved the ssRNA substrate under all tested conditions. In contrast, SisCsx1 showed little activity towards a dsRNA molecule generated by annealing SS1–40 with SS1–40T, the complementary RNA (Figure 1D). These results indicated that SisCsx1 is a metal-independent ssRNA-specific ribonuclease.

**RNAs carrying a poly(A) tail strongly elevates SisCsx1 RNase activity**

Two additional ssRNA substrates (SS1–46 and S10) were tested for SisCsx1 cleavage to further study specificity of RNA cleavage by the enzyme, both of which contain a 6 nt poly(A) tail that is absent from SS1–40 (Supplementary Table S2). Each RNA substrate was incubated with SisCsx1 over time during which samples were taken for RNA cleavage analysis by denaturing gel electrophoresis. Surprisingly, we found that SisCsx1 cleaved about 95% of SS1–46 and S10 in 60 min (Supplementary Figure S2), whereas ~70% of SS1–40 was cleaved under the same conditions, suggesting that SisCsx1 could be more active on poly(A) RNA molecules, or poly(A) RNAs could function as an activator to the RNase.

To investigate which scenarios could be true, poly(A)-containing SS1–46 and poly(A)-less SS1–40 were radio-labeled individually, and each labeled substrate was then mixed with 10-fold excess unlabeled SS1–46 or SS1–40 RNA substrate. The mixtures were subsequently used as RNA substrates to examine the cleavage activity of SisCsx1. We found that, while addition of 10-fold excess SS1–40 did not influence the cleavage of labeled SS1–46 or SS1–40, addition of 10-fold excess SS1–46 significantly enhanced the cleavage of both labeled SS1–40 and SS1–46 (Supplementary Figure S3). These results suggested that the poly(A) tail
present in SS1–46 could function as an activator to the ribonuclease activity of SisCsx1.

The hypothesis was further tested with 5′-S1A7–3′ (S−C or G, Supplementary Table S2), a 10-nt RNA oligo containing seven adenosines at the 3′ end. As shown in Figure 2A, in the presence of 200 nM S1A7 and a 15-fold lower amount of SisCsx1 (200 nM), all labeled SS1–40 or SS1 RNA (25 nM) was cleaved within 20 min of incubation. In comparison, the same amount of the enzyme exhibited little activity on SS1–40 in the absence of S1A7. To test if the activation could be poly(A) RNA-specific, several other short RNAs and DNAs were synthesized, including A7S3 (seven A at the 5′-end and three C/G at the 3′-end), poly(C) (C10), poly(G) (CG9), poly(U) (U10), poly(A) DNA (dA40), poly(T) DNA (dT40), poly(C) DNA (dC40), poly(G) DNA (dG40) and dsDNA (dA40+dT40) and tested for their effects on the Csx1 activity. As shown in Figure 2B and Supplementary Figure S4A, only RNAs containing a poly(A) sequence were capable of activating the SisCsx1 RNase, neither DNAs nor any of the other types of RNA showed any stimulation of the RNase activity. Further, SisCsx1 exhibited higher activity in the presence of S1A7 of high concentrations (Supplementary Figure S4B). Together, these results indicated that the RNase activity of SisCsx1 is specifically activated by poly(A) RNAs.

To investigate if the activation of SisCsx1 requires constant presence of the activator, we attempted to separate SisCsx1 from a pre-incubated mixture of SisCsx1 and S1A7 by gel filtration chromatography, and analysis of the cleavage activity of protein fractions by incubation with labeled SS1–40 substrate showed that the re-purified SisCsx1 was not active unless the fraction containing S3A7 was added again (Figure 2C and D), indicating that depletion of the poly(A) RNA activator inactivated the enzyme.

Activated SisCsx1 is complementary to the Cmr-α effector complex on RNA cleavage

Next, we analyzed the substrate turnover rate of activated SisCsx1 by supplementing increasing concentrations of unlabeled SS1–40 into the reaction mixture. As shown in Figure 3A and B, the cleavage of SS1–40 exhibited a nearly

---

**Figure 1.** SisCsx1 is a metal-independent single stranded-specific ribonuclease. (A) SDS-PAGE analysis of the purified SisCsx1. (B) MALS estimation of the size of SisCsx1 in solution. (C) The cleavage activity of SisCsx1 is independent of metal ions. Labeled SS1–40 RNA was incubated with 3 μM SisCsx1 for 1 h in the presence of metal ions or EDTA as indicated, followed by separation by denaturing gel electrophoresis. (D) ssRNA (SS1–40) and dsRNA (SS1–40+SS1–40T) were incubated with 3 μM SisCsx1 for 40 min and then analyzed by denaturing gel electrophoresis.
Figure 2. SisCxs1 is activated by poly(A) RNA. (A) Labeled SS1–40 and S10 (about 20 nM) were incubated with SisCxs1 at indicated concentration with or without 200 nM S3A7 for 20 min. Then, the reactions were stopped and analyzed by denaturing gel electrophoresis. (B) Labeled SS1–40 was incubated with 100 nM SisCxs1 in the presence of 200 nM indicated RNA oligos for 20 min, followed by separation by denaturing gel electrophoresis. (C) Separation of SisCxs1 and S3A7 by gel filtration chromatography. SisCxs1 (panel I), S3A7 (panel II) and a mixture of pre-incubated SisCxs1 and S3A7 (panel III) were loaded onto a Superdex 200 10/300 GL analytical column, and the UV absorbance at 280 nm and 254 nm was shown. (D) Cleavage of SS1–40 by the fractions from the panel III of (C). Labeled SS1–40 was incubated with about 200 nM SisCxs1 from 10-ml fraction or about 200 nM SisCxs1 from 10-ml fraction plus about 200 nM S3A7 from 19-ml fraction for 10 min and resolved by denaturing gel electrophoresis.

The HEPN domain is responsible for basal and activated RNA cleavage activity

We showed that SisCxs1 exhibits ribonuclease activity at two distinct levels, i.e. the basal level and the activated level in the presence of an RNA activator. Previous studies revealed that a conserved motif, R-X4–6-H, located in the C-terminal HEPN domain is important for the ribonuclease activity of Csx1 and Csm6 (11,27). Here, we investigated whether the SisCxs1 HEPN domain could also be responsible for the ribonuclease activity. Two alanine mutations were introduced to the well-conserved arginine and histidine residues in the R-X4–6-H motif of the HEPN domain, giving Csxs1M1 (Supplementary Figures S6 and S7). The resulting mutant gene was expressed in S. islandicus as recombinant protein and assayed for RNA cleavage. The results showed that Csxs1M1 became inactive in RNA hydrolysis, even at a relatively high concentration (1 μM) or in the presence of the activator S3A7 (Figure 4A), indicating that both the basal and the activated ribonuclease activity of SisCxs1 resides in the HEPN domain.

Next, substrate specificity of the enzyme was investigated in the presence of the S3A7 activator with different forms of nucleic acids, including ssRNA, dsRNA, ssDNA and dsDNA. As in the absence of any activator, SisCxs1 only de-

linear response to increasing substrate concentration until 5 μM, when the molar concentration of substrate is 25 times of SisCxs1 and S3A7, indicating a fast turnover rate of SisCxs1 on substrate. This is in contrast to the low turnover rate on target RNA of the type III-B SisCmr-α-SS1 complex (24) where the enzyme was saturated by target RNA of doubled molar concentration (Supplementary Figure S5).

Then, we analyzed the cleavage pattern by mixed SisCmr-α-SS1 and SisCxs1. SS1–40 of different concentrations (25 nM and 5 μM) was cleaved by a mixture of 25 nM SisCmr-α-SS1 and 100 nM SisCxs1 with 500 nM S3A7. Separate cleavage by SisCmr-α-SS1 and SisCxs1 with S3A7 was also analyzed as references. As shown in Figure 3C, SisCmr-α-SS1 efficiently cleaved the equivalent SS1–40 but hardly degraded 5 μM SS1–40, while SisCxs1 efficiently degraded both 25 nM and 5 μM SS1–40 at almost equal rate. In the presence of both RNases, 25 nM SS1–40 was mostly cleaved by SisCmr-α-SS1 as judged from the distribution of the products, while 5 μM SS1–40 was cleaved by SisCxs1. The data indicate that SisCxs1 did not affect the cleavage of equivalent target RNA by Cmr-α, but efficiently degraded RNA substrate beyond the cleavage capability of Cmr-α, suggesting that SisCxs1 plays a complementary role to Cmr-α.
grated ssRNA (Figure 4B), indicating that poly(A) RNA activator does not have any influence on substrate specificity of the enzyme.

Previous studies showed that Csx1/Csm6 yield products carrying a 2',3'-cyclic phosphate or 3'-phosphate and 5'-OH group (11,27). To analyze whether SisCsx1 could possess the same cleavage property, the 3'-end group of the cleavage products generated by SisCsx1 was analyzed using poly(A) polymerase (PAP) extension assay as reported previously (27) since PAP is capable to add poly(A) tail to any RNA that carries the 3'-OH group. 5'-labeled SS1–40 was cleaved by SisCsx1, and the resulting cleavage products were treated with PAP and ATP for 10 min. Analysis of RNAs in the reactions by denaturing gel electrophoresis showed that, while uncleaved substrate was extended by PAP, none of the cleavage products were extended (Figure 4C), indicating that all the products lack the 3'-OH group. Therefore, cleavage products generated by SisCsx1 should carry the same end groups as for the products generated by other known HEPN domain nucleases. Further, SisCsx1 produced the same cleavage products in the presence or absence of S1A7 RNA activator, indicating that the RNA did not influence the cleavage pattern of the enzyme.

Activation of SisCsx1 RNase requires an RNA carrying 4 nt poly(A) tail

To determine the minimal requirement of an RNA activator for SisCsx1, four additional 10 nt RNAs were synthesized, i.e. S8A2, S7A3, S6A4 and S5A5. Investigation of their capability to facilitate SisCsx1 RNase activity showed that while the enzyme failed to cleave SS1–40 substrate in the presence of S8A2 and S7A3 (Figure 5A and B), S6A4, S5A5 and S1A7 strongly activated SS1–40 cleavage by SisCsx1, and furthermore, the activation was equally efficient by the three RNA activators carrying four, five or seven adenosines at the 3'-end (Figure 5A and B). The results indicated that the minimal size of poly(A) in an RNA activator is 4 nt and
the presence of additional adenosine nucleotides (>4 nt) in RNA activators does not further enhance the activation.

To address whether additional 3′-end non-adenosine nucleotides would affect SisCsx1 activation, five additional RNA oligos were synthesized, i.e. CA4 carrying an 3′-OH group, CA4-p containing a 3′ phosphate group, and CA4S1, CA4S3 and CA4S6 that carry one, three or six additional C or G nucleotides at the 3′-end, respectively. Analysis of SisCsx1 activity in the presence of each of these RNAs showed that addition of one or more nucleotides at 3′-end significantly impaired the activation, indicating that SisCsx1 activation requires a 3′-poly(A) tail (Figure 5C).

Then, modifications of poly(A) in CA4 were designed to yield an additional insight into the activation mechanisms, including 4 variants lacking 2′-hydroxyl group at different positions (CA3-dA, CA2-dA-A, CA-dA-A2, C-dA-A3), one phosphorothioate backbone modification in which a non-bridging oxygen of the sugar-phosphate linkage is replaced with a sulfur atom between each two of the four adenosines (C*A4). Each of these RNAs was tested for SisCsx1 RNase activation, and we found that none of them were capable of activating the SisCsx1 RNase (Figure 5C). Together, the results indicated that the integrity of 3′-tetra-rA is essential for the activation.

RNA activators bind to SisCsx1 as a ligand

The activation of SisCsx1 by 3′-poly(A) RNA suggested that SisCsx1 could bind to the poly(A) tail of a RNA molecule as ligand to allosterically regulate the RNase activity. To test the hypothesis, the electrophoretic mobility shift assay (EMSA) was employed to detect the formation of RNA-protein complexes. As shown in Supplementary Figure S8, the labeled S3A7 did not show any retardation in migration in the presence of excess SisCsx1 (4.5 μM), indicating that the interaction between SisCsx1 and the RNA ligand could be weak. Then, UV cross-link assay was employed to investigate the interaction of SisCsx1 and 3′-poly(A) RNA. SisCsx1 was incubated with increasing concentrations of unlabeled S3A7 and exposed to UV irradiation, and the samples were then analyzed by SDS-PAGE. The results showed that in the absence of any RNA ligand, UV irradiation yielded additional protein bands that appeared either larger or smaller than the original SisCsx1 protein, and in the presence of an excess amount of RNA ligand S3A7 (50 μM), UV radiation produced a new band corresponding to the size of SisCsx1 dimer, suggesting that two SisCsx1 molecules were crosslinked together in the presence of the S3A7 ligand (Figure 6A).

We further compared the interaction of SisCsx1 with S3A7 and a few non-activator RNAs by UV crosslink. Labeled RNA oligos were incubated with SisCsx1, followed by UV irradiation. Then, the samples were analyzed by SDS-PAGE and visualized by phosphorimaging and Coomassie blue staining. The results showed that S3A7 was associated with SisCsx1 in the SDS-polyacrylamide gel, while a much lower level of CA10, CG10, U10, S7A3 and A7S3 were observed to associate with SisCsx1, suggesting that SisCsx1 specifically binds to 3′-end poly(A) RNA (Figure 6B). These results were in good agreement with those obtained from their RNase assay.

Next, we analyzed the affinity of CA4 and its variants to SisCsx1, and this revealed that addition of one or more nucleotides at 3′-end significantly reduced the affinity to SisCsx1, in agreement with their effects on SisCsx1 activity (Figure 6C). However, removal of 2′-hydroxyl group of each adenosine or phosphorothioate modification of the RNA ligand greatly weakened or abolished the activation of SisCsx1 although these RNAs showed an elevated affinity to the protein (Figure 6C). Taken together, the data indicated that binding to RNA ligand is necessary but not sufficient for SisCsx1 activation.
Figure 5. Determination of the activator for SisCx1. (A) Activation of SisCx1 requires four adenosines. Labeled SS1–40 was incubated with 100 nM SisCx1 with 2 μM S6A3, S7A3, or 200 nM S6A4, S5A5 and S3A7 or without any short RNA oligo, followed by denaturing gel electrophoresis. (B) Quantification of the percentage of cleaved substrate from (A). ‘None’ means no 10-nt RNA oligo. Error bar represents S.D. of three independent experiments. (C) Labeled SS1–40 was incubated with 200 nM SisCx1 with 200 nM CA4 and its variants (Supplementary Table S2) for 20 min and analyzed by denaturing gel electrophoresis.

The CARF domain is responsible for ligand binding

The CARF domain of Csx1 contains a predicted ligand-binding site (12). It was reported that Csx3, a single domain protein of the CARF superfamily, co-crystallizes with tetranucleotides (9,29). To investigate if SisCx1 CARF domain could be responsible for binding of poly(A) RNA and for regulation of the RNase activity in the HEPN domain, we expressed five Csx1 mutants carrying mutations in the conserved amino acids of the CARF domain (Supplementary Figures S6 and S7), i.e. Csx1-M2 (G95L-A97L-A99L), M3 (D50L-S51L), M4 (G9L), M5 (G156L) and M6 (H155L). These mutant proteins were expressed and purified from E. coli (Supplementary Figure S7) since attempts to produce them from S. islandicus failed. These Csx1 proteins were individually incubated with label SS1–40 substrate in the absence of S3A7 to analyze their basal ribonuclease activity, and this revealed that all CARF mutants possessed an RNA cleavage activity comparable to the WT enzyme at the concentration of 1 μM (Figure 7A). However, when reduced to 200 nM, none of the mutant enzymes exhibited a detectable cleavage activity in the presence of S3A7, suggesting that they all lost the capability to be activated by poly(A) RNA (Figure 7B). Further, UV crosslink assay revealed that, while M4 and M5 exhibited a lower affinity to S3A7 compared to WT, the binding of S3A7 to M2, M3 and M6 went to an undetectable level (Figure 7C). In contrast, HEPN mutation did not affect the affinity of the mutant enzyme to S3A7 (Supplementary Figure S9). Taken together, the data indicated that the conserved amino acids in the CARF domain are responsible for ligand binding and/or allosteric regulation to activate the RNase activity of the HEPN domain.

DISCUSSION

In this study, we have characterized the S. islandicus Csx1, a type III-B-associated Cas accessory protein that contains a CARF and a HEPN domain. Functional analyses of the two domains have revealed that the CARF domain can bind to an RNA ligand carrying a tetraadenylate tail and that the ligand binding activates the RNA cleavage by the HEPN domain.
The tetraadenylate RNA ligand was discovered accidentally. When two categories of RNA, which were used to demonstrate the RNA-activated DNA cleavage by Cmr-α in a previous work (24), were employed as substrate to investigate the SisCxl RNase activity, i.e. the cognate target RNAs carrying a 6 nt 3′-poly(A) tail in the corresponding position of the repeat tag (SS1–46 and S10) and the repeat tag-less target RNA (SS1–40), we found that SisCxl exhibits a higher activity to RNAs containing a poly(A) tail compared with non-poly(A) RNAs, and this suggests that poly(A) RNAs can strongly activate the basal RNase activity of SisCxl. This hypothesis was then tested with a series of small RNAs, and this has revealed that an activator RNA should carry at least 4-nt adenosines at the 3′-tail. We further show that SisCxl directly binds to the RNA ligand. The ligand-enzyme interaction is dependent on the CARF domain of SisCxl, and the ligand-binding is required to activate the RNA cleavage activity by the HEPN domain.
Moreover, poly(A) segment at 5’-end or in the middle of small RNAs exhibits much lower affinity to SisCsx1 and significantly reduced capability to stimulate the enzyme, suggesting that the RNA ligand could be recognized by the enzyme from 3’-end. In addition, the 2’-OH groups of poly(A) adenosines are not important for ligand binding but essential for enzyme activation, revealing additional insight to the mechanisms of the allosteric regulation of SisCsx1 activity.

We have also revealed that the interaction between the poly(A) RNA ligand and SisCsx1 is rather dynamic and dissociation of RNA ligands from the enzyme can result in an almost inactive enzyme. The findings imply that SisCsx1 could be activated by RNA polyadenylation and inactivated by degradation of 3’-poly(A) RNAs. Interestingly, this is consistent with the RNA decay pathway present in most prokaryotes, which functions in two consecutive steps: addition of 3’-poly(A) tail onto RNA (RNA polyadenylation) and degradation polyadenylated RNA (30–32). To this end, it is tempting to assume that poly(A) RNAs synthesized by exosomes in *Sulfolobus* readily activates Csx1. Since tetraadenylate-Csx1 interaction is dynamic, this renders it possible for the archaeon to readily activate the enzyme when polyadenylated RNAs are abundant and to readily inactivate the enzyme when poly(A) RNAs are degraded. Strikingly, recent study on the regulation of RNA polyadenylation in *S. solfataricus* has shown that DnaG and SmAP proteins enhance RNA polyadenylation activity (33,34), and these data provide not only the first examples of regulation of RNA polyadenylation in response to cellular stresses, but also a mechanism of allosteric regulation of Csx1 by poly(A) RNAs in vivo.

Furthermore, transcripts in prokaryotes end up with U-rich tail (35,36), which could avoid activation of Csx1 at normal growing conditions. In addition, the widespread CARF domain toxins, i.e. CARF domain proteins carrying a C-terminal nuclease domain (6), are absent in halobacteria (Supplementary Table S3), an archaeal class that lack RNA polyadenylation (32), providing an indirect evidence supporting that regulation of RNA polyadenylation in response to cellular stresses, but also a mechanism of allosteric regulation of Csx1 by poly(A) RNAs in vivo.

Most recently, it has been reported that type III-A effector Csm complex synthesizes a cyclic oligonucleotide upon binding to target RNA and that the cyclic oligonucleotide activates the ribonuclease activity of type III-A associated Csm6 protein (37,38). The Csm6 activation is strictly dependent on an active type III-A effector complex, the presence of a spacer matching viral genome and the transcription of the protospacer (37). Therefore, the response is highly specific. However, in another report, RNA cleavage activity by Csm6 was found to be essential for the type III-A-mediated antiviral immunity when the protospacer is from a late-expressed gene or when the protospacer is mutated (26). Since the Csm effector complex cannot synthesize the cyclic oligonucleotide required for Csm6 activation either in a timely fashion or effectively under those conditions (37), this suggests that an alternative mechanism is required for Csm6 activation to yield the observed results. In this respect, another RNA ligand such as poly(A) could have played the role instead of the cyclic oligonucleotide. Indeed, it has been shown that the RNase activity of Csm6 can also be strongly activated by poly(A) RNAs (38). Together, these data suggest that the tetraadenylate activation of CARF-domain RNases is evolutionarily conserved in Bacteria and Archaea although it remains to be demonstrated whether the cyclic oligoadenylate synthesis is conserved in other type III CRISPR–Cas systems.

On the other hand, genes coding for CARF domain nucleases are dispersed in bacterial and archaeal genomes, a number of which are not associated with type III CRISPR–Cas systems (6,10). Our finding that poly(A) RNAs can function as an activator to the *S. islandicus* SisCsx1 can be generally applied to all CARF domain nucleases regardless whether they are associated with any CRISPR–Cas systems. Thus, RNA polyadenylation-dependent activation of CARF domain nucleases presents an alternative mechanism of enzyme activation.

Indeed, SisCsx1 cleaves unspecific ssRNA substrates with a high turnover rate, which is in contrast to the low turnover RNA cleavage by Cmr-α. Therefore, SisCsx1 can have more important roles in destruction of viral transcripts to limit the infectious cycle, relative to the III-B Cmr-α system. Furthermore, the rapid destruction of ssRNAs by activated SisCsx1 exhibits a good potential to degrade large amounts of host mRNAs, leading to growth inhibition, cell dormancy or even cell death of the host as well as to prevention of virus replication and spreading (39). In eukaryotes, it has been shown that a ubiquitous RNase, RNase L is activated by 2’-5’ oligoadenylate upon virus infection and the enzyme is responsible for degradation both viral and host RNA and induce autophagy and apoptosis (40,41). Although it remains to be demonstrated, several prokaryotic RNases could play a similar role, and these include Csm6 proteins from several bacterial species (37,38) and their distant homolog SisCsx1 from a hyperthermophilic archaeon reported here, and probably all identified Csx1/Csm6 homologs (6). Therefore, allosteric regulation of RNase activity may represent a general strategy for antiviral defense in all three domains of life.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

We thank members of both laboratories for stimulating discussions.

**FUNDING**

Danish Council for Independent Research [DFF-0602-02196, DFF-1323-00330]; The Augustinus Foundation [15-3934]; Brodrene Hartmanns Foundation [A29992]; Natural Science Foundation of China [31128011]; Scientific and Technological Self-Innovation Foundation of Huazhong Agricultural University [2014RC011 to Q.S.]. The Novo Nordisk Foundation Center for Protein Research is supported financially by the Novo Nordisk Foundation [NNF14CC0001]; Danish Cancer Society (to G.M.). Funding for open access charge: University of Copenhagen.
Conflict of interest statement. None declared.

REFERENCES