Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Crystal structure and CRISPR RNA-binding site of the Cmr1 subunit of the Cmr interference complex

Jiali Sun,[a]‡ Jae-Hyun Jeon,[b]‡ Minsang Shin,[a]‡ Ho-Chul Shin,[b] Byung-Ha Oh[b]* and Jeong-Sun Kim[a]*

[a]Department of Chemistry, Chonnam National University, Gwangju 500-757, Republic of Korea, and [b]Department of Biological Sciences, KAIST Institute for the Biocentury, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Republic of Korea

‡ These authors contributed equally to this work.

Correspondence e-mail: bhoh@kaist.ac.kr, jsunkim@chonnam.ac.kr

Received 1 July 2013
Accepted 5 November 2013

PDB reference: *Af* Cmr1, 4l6u

A multi-subunit ribonucleoprotein complex termed the Cmr RNA-silencing complex recognizes and destroys viral RNA in the CRISPR-mediated immune defence mechanism in many prokaryotes using an as yet unclear mechanism. In *Archaeoglobus fulgidus*, this complex consists of six subunits, Cmr1–Cmr6. Here, the crystal structure of Cmr1 from *A. fulgidus* is reported, revealing that the protein is composed of two tightly associated ferredoxin-like domains. The domain located at the N-terminus is structurally most similar to the N-terminal ferredoxin-like domain of the CRISPR RNA-processing enzyme Cas6 from *Pyrococcus furiosus*. An ensuing mutational analysis identified a highly conserved basic surface patch that binds single-stranded nucleic acids specifically, including the mature CRISPR RNA, but in a sequence-independent manner. In addition, this subunit was found to cleave single-stranded RNA. Together, these studies elucidate the structure and the catalytic activity of the Cmr1 subunit.

1. Introduction

Many bacteria and archaea have an inheritable RNA-based immune defence mechanism for destroying invading phages and plasmids. Central to this mechanism is the presence of clustered regularly interspaced short palindromic repeats (CRISPRs) and *CRISPR-associated* (*cas*) genes in prokaryotic genomes. CRISPR loci contain two alternating short DNA sequences, which are called repeats and spacers. Whereas the repeats are composed of 25–50 base pairs and are nearly identical to each other, the spacers are composed of 21–72 base pairs and their sequences vary (Godde & Bickerton, 2006; Grissa *et al.*, 2007*b*; Garneau *et al.*, 2010). Bacteria and archaea acquire the spacer sequences from foreign DNA upon viral and plasmid invasion (Bolotin *et al.*, 2005; Mojica *et al.*, 2005; Pourcel *et al.*, 2005). *Cas* genes are tandem arrays of genes which are typically located in the vicinities of CRISPR loci and they encode proteins that mediate the RNA-based defence processes. Initially, four *cas* genes (*cas*1–*cas*4) were identified that are commonly present in CRISPR-containing prokaryotic genomes (Jansen *et al.*, 2002). A wide variety of *cas* genes have subsequently been identified, and the CRISPR–Cas systems have been classified into three major types that are further divided into several subtypes and a few variants (Makarova *et al.*, 2011).

The CRISPR-mediated immune defence process is usually divided into three stages (Deveau *et al.*, 2010). In the first stage, a short DNA segment termed a protospacer is acquired from an invading phage or plasmid and is inserted next to the leader sequence, which is located at the 5′ end of a CRISPR locus (Barrangou *et al.*, 2007; Garneau *et al.*, 2010). While the

underlying molecular mechanisms are elusive, Csn2 (at least in *Streptococcus thermophilus*), the universally present Cas1 and host DNA recombination and repair factors have been implicated as being involved in this process (Barrangou *et al.*, 2007; Makarova *et al.*, 2006; Babu *et al.*, 2011; Wiedenheft *et al.*, 2009). In the second stage, small mature CRISPR RNAs termed crRNAs are generated from a long transcript of a CRISPR locus (Haurwitz *et al.*, 2010; Deltcheva *et al.*, 2011; Carte *et al.*, 2008; Wang *et al.*, 2011). In the type I system, crRNA biogenesis requires Cas6e in *Escherichia coli* and Cas6f in *Pseudomonas aeruginosa*. These two Cas proteins, as a subunit of the CASCADE (CRISPR-associated complex for antiviral defence) complex, are CRISPR-specific endonucleases that recognize the sequence and shape of the pre-CRISPR RNA. They cleave the repeat sequence eight nucleotides upstream of the spacer sequence, resulting in crRNAs with a spacer flanked by an eight-nucleotide repeat segment at the 5′ end and a remaining repeat segment at the 3′ end (Gesner *et al.*, 2011; Haurwitz *et al.*, 2010; Sashital *et al.*, 2011). In the type II system, as found in *S. pyogenes*, a completely different crRNA biogenesis mechanism operates, which involves Cas9 and a host RNase III that cleaves the double-stranded region formed by pre-crRNA and a short RNA transcript complementary to the repeat sequence known as a *trans*-activating CRISPR RNA (Deltcheva *et al.*, 2011). In the type III system, as found in *Pyrococcus furiosus*, Cas6 recognizes CRISPR RNA differently, but it also cleaves the cognate RNA eight nucleotides upstream of the spacer sequence (Wang *et al.*, 2011). In the final stage, which is referred to as the interference stage, a crRNA-containing ribonucleoprotein complex cleaves non-self invasive genetic elements (either DNA or RNA) when the crRNA spacer binds to the complementary target sequence (protospacer; Brouns *et al.*, 2008; Hale *et al.*, 2009). To date, three different effector complexes have been identified: CASCADE, which targets DNA (in type I; Jore *et al.*, 2011; Nam *et al.*, 2012), Cas9, which targets DNA (in type II; Gasiunas *et al.*, 2012), and the Csm complex (in type III-A) and a related Cmr (Cas module RAMP) complex (in type III-B), which target DNA and RNA, respectively (Hale *et al.*, 2009; Zhang *et al.*, 2010).

The Cmr interference complex in *Sulfolobus solfataricus* is composed of seven subunits (Cmr1–Cmr7). It cleaves target RNAs in a sequence-specific manner (rather than using a ruler mechanism) and requires the eight nucleotide repeat-derived 5′ sequence in a guide crRNA (Zhang *et al.*, 2012). The crystal structure of Cmr7 and an electron-microscopic study revealed a crab claw-like structure of the holocomplex (Zhang *et al.*, 2012), which is distinct from the sea horse-like structure of the *E. coli* CASCADE (Jore *et al.*, 2011). In comparison, the RNA-silencing Cmr complex of *P. furiosus* is composed of six Cmr proteins (Cmr1–Cmr6; Hale *et al.*, 2009). A 39- or 45-nucleotide crRNA is an integral part of this complex and guides the cleavage of target RNAs using a 3′ molecular ruler mechanism (Hale *et al.*, 2012). In contrast, a 5′ ruler mechanism has recently been suggested for the *Thermus thermophilus* Cmr holocomplex based on the pattern of target RNA cleavage (Staals *et al.*, 2013).

The crystal structure of a subcomplex between Cmr2 and Cmr3 has been determined (Shao *et al.*, 2013), and electron-microscopic (EM) structures of the Cmr holocomplexes of *P. furiosus* and *T. thermophilus* and the Csm complex of *S. solfataricus* have recently been reported (Spilman *et al.*, 2013; Staals *et al.*, 2013; Zhang *et al.*, 2012). The overall structures of these complexes are strikingly similar to that of CASCADE. Despite this significant progress, it is as yet unknown which subunit is responsible for the slicing activity. Previously, Cmr2 (also known as Cas10), the largest subunit of the complex, was predicted to be the RNA-cleaving nuclease because it contains an N-terminal HD phosphohydrolase domain. However, recent structural and biochemical studies of Cmr2 demonstrated that neither the HD domain nor the two newly identified adenylyl cyclase-like domains in this protein are required for the catalytic function of the silencing complex (Cocozaki *et al.*, 2012; Zhu & Ye, 2012). Based on the observation of multiple cleavages of a target RNA by a *T. thermophilus* Cmr holocomplex which contains four copies of Cmr4, the Cmr4 subunit was suggested to be a ribonuclease (Staals *et al.*, 2013). Cmr1 or Cmr6 was also suspected to be a ribonuclease subunit which might be responsible for one of the multiple cleavages (Staals *et al.*, 2013).

The hyperthermophilic archaeon *Archaeoglobis fulgidus* encodes six Cmr proteins that are highly homologous to the *P. furiosus* counterparts that constitute the Cmr interference complex. Here, we report the crystal structure of Cmr1 (AF1868) of *A. fulgidus*, referred to as *Af* Cmr1 in the following, and the identification of this subunit as a nuclease specific for single-stranded nucleic acids.

2. Materials and methods

2.1. Cloning and protein production

The full-length *Af* Cmr1 gene was amplified by polymerase chain reaction (PCR) from the chromosomal DNA of *A. fulgidus* with primers designed for ligation-independent cloning (Aslanidis & de Jong, 1990). The PCR product was treated with T4 DNA polymerase (New England Biolabs) and inserted into a vector derived from the pET-21a plasmid (Novagen). This vector was designed to express the cloned gene fused to maltose-binding protein (MBP)-His_6 and the *Tobacco etch virus* (TEV) cleavage sequence at the N-terminus. *E. coli* BL21(DE3) Star strain transformed with the expression construct was grown in Luria–Bertani medium. After induction by adding 0.5 m*M* IPTG, the culture medium was maintained for 8 h at 37°C. The harvested cells were resuspended and disrupted by ultrasonication in a buffer solution consisting of 20 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl (buffer *A*). The supernatant was loaded onto a 5 ml HisTrap chelating column (GE Healthcare). The column was extensively washed with buffer *A* and the bound proteins were eluted with a linear gradient from 0 to 500 m*M* imidazole in buffer *A*. The eluted sample was dialyzed against a buffer solution consisting of 20 m*M* Tris–HCl pH 7.5, 100 m*M* NaCl (buffer *B*), and the MBP-His_6 tag was cleaved with TEV. *Af*

Table 1
Data-collection and structure-refinement statistics.

Values in parentheses are for the highest resolution shell.

	Native	SeMet derivative
Data collection		
Space group	$P2_1$	$P2_1$
Unit-cell parameters (Å, °)	a = 70.7, b = 64.16, c = 79.25, $\alpha = \gamma$ = 90.0, β = 93.74	a = 71.71, b = 64.43, c = 80.13, $\alpha = \gamma$ = 90.0, β = 94.13
Wavelength (Å)	1.1159	0.9798
Resolution (Å)	79.1–2.50 (2.64–2.50)	71.5–2.65 (2.79–2.65)
$R_{\rm merge}$ (%)	5.9 (68.6)	6.4 (60.3)
$\langle I/\sigma(I)\rangle$†	9.6 (1.4)	9.3 (1.5)
Completeness (%)	99.1 (98.8)	98.5 (98.1)
Multiplicity	3.6 (3.7)	3.6 (3.7)
Figure of merit (*SOLVE*/*RESOLVE*)		0.22/0.56
Refinement		
Resolution (Å)	79.1–2.5	
No. of reflections	24479	
$R_{\rm work}/R_{\rm free}$ (%)	23.0/25.3	
No. of atoms		
Protein	4546	
Water	51	
R.m.s. deviations		
Bond lengths (Å)	0.003	
Angles (°)	0.70	
Average B values (Å^2)		
Protein	37.5	
Water	50.9	
Ramachandran plot (%)		
Favoured	95.6	
Allowed	4.4	
Outliers	0	

† $\langle I/\sigma(I)\rangle$ reaches 2.0 at 2.65 Å for the native data set and at 2.8 Å for the the SeMet-derivative data set.

Cmr1 was isolated using a 5 ml HiTrap SP cation-exchange column (GE Healthcare) with a linear gradient from 0 to 1 *M* NaCl in buffer *B*. Seleno-L-methionine (SeMet)-substituted *Af* Cmr1 was prepared similarly to the native protein. The purified protein was concentrated to 10 mg ml^{-1} after a buffer change to 20 m*M* Tris–HCl pH 8.5, 200 m*M* NaCl.

2.2. Crystallization, data collection and structure determination

Crystallization of *Af* Cmr1 was attempted at 22°C using the sitting-drop vapour-diffusion method. The initial crystals were obtained from a precipitant solution consisting of 0.1 *M* ammonium acetate, 0.1 *M* bis-tris pH 5.5, 17%(*w*/*v*) polyethylene glycol 10 000. This condition was optimized by a grid search in 24-well Linbro plates using the hanging-drop vapour-diffusion method at 22°C, in which 1 µl protein sample and 1 µl precipitant were mixed together and equilibrated against 0.2 ml precipitant. Suitable crystals for diffraction experiments were obtained using a precipitant consisting of 20%(*w*/*v*) polyethylene glycol 6000, 0.1 *M* NaCl, 0.1 *M* bis-tris pH 5.5, 60 m*M* β-mercaptoethanol. SeMet-substituted protein crystals were obtained using the same crystallization conditions. For diffraction experiments, crystals were briefly immersed into the precipitant solution containing an additional 10%(*v*/*v*) glycerol as a cryoprotectant and were immediately placed in a 100 K nitrogen-gas stream. Using a SeMet-labelled *Af* Cmr1 crystal, single-wavelength anomalous diffraction (SAD) data were collected at the Se peak wavelength (0.9798 Å) on beamline 8.3.1 at the Advanced Light Source with an oscillation of 1° per frame, an exposure of 5 s per frame and a crystal-to-detector distance of 350 mm. A native data set was collected at a wavelength of 1.1159 Å on the same beamline with an oscillation of 1° per frame, an exposure of 3 s per frame and a crystal-to-detector distance of 300 mm. In both cases a total of 180 images were collected using an ADSC Q315r CCD detector. The mosaicities of the SeMet-labelled and native crystals were 0.73 and 0.63°, respectively. Indexing, integration and scaling of the reflections were conducted using *MOSFLM* (Leslie, 1992) and the programs implemented in the *ELVES* software suite (Holton & Alber, 2004). Eight of the expected 14 Se sites in the asymmetric unit were identified at a resolution of 2.65 Å using *PHENIX* (Adams *et al.*, 2010) combined with *SOLVE* (Terwilliger & Berendzen, 1999). Electron-density modification was performed using *PHENIX* (Adams *et al.*, 2010) combined with *RESOLVE* (Terwilliger, 2003), resulting in the automatic modelling of approximately 40% of the residues. Further model building was performed manually using *WinCoot* (Emsley & Cowtan, 2004) and subsequent refinement was performed with *PHENIX* (Adams *et al.*, 2010). Noncrystallographic symmetry was not used in the structure refinement, while the asymmetric unit of the crystal contained two *Af* Cmr1 molecules. The conformations of the two molecules were virtually identical (r.m.s.d. of 0.82 Å for all C$^\alpha$ atoms). The data-collection and refinement statistics are summarized in Table 1.

2.3. Production of *Af* Cmr1 mutants

Charge-inversion mutants of *Af* Cmr1 were generated using the overlapping PCR protocol and the mutations were confirmed by DNA sequencing. The primer sequences used for site-directed mutagenesis were 5′-C AAA GCC GAA ATC **GAA** GCT GCA TCG A-3′ and 5′-T CGA TGC AGC **TTC** GAT TTC GGC TTT G-3′ (R29E mutation), 5′-T AAA GGC TTG ATG **GAA** TGG TGG TTC AGG G-3′ and 5′-C CCT GAA CCA CCA **TTC** CAT CAA GCC TTT A-3′ (R38E mutation), 5′-GG TGG TGG TTC **GAA** GCT CTG TCC GG-3′ and 5′-CC GGA CAG AGC **TTC** GAA CCA CCA CC-3′ (R42E mutation), 5′-T GGG ATT GGT TTT **GAA** TGC TCT CGT GGA G-3′ and 5′-C TCC ACG AGA GCA **TTC** AAA ACC AAT CCC A-3′ (R145E mutation), 5′-T TTT AGA TGC TCT **GAA** GGA GCG GGG TCA C-3′ and 5′-G TGA CCC CGC TCC **TTC** AGA GCA TCT AAA A-3′ (R148E mutation), 5′-G AGA GGT ACA AAA AAA GAC **GAA** AGA GCA TCT CCT ATT AAA-3′ and 5′-TTT AAT AGG AGA TGC TCT **TTC** GTC TTT TTT TGT ACC TCT C-3′ (R274E mutation), and 5′-A GGT ACA AAA AAA GAC AGG **GAA** GCA TCT CCT ATT AAA GTC-3′ and 5′-GAC TTT AAT AGG AGA TGC **TTC** CCT GTC TTT TTT TGT ACC T-3′ (R275E mutation). Double mutants were produced using the same method, except that the R38E and R42E mutant genes were used as the template for PCR. The

Table 2
Oligonucleotides used in this study.

The repeat sequences in the CRISPR loci of *A. fulgidus* are indicated in bold.

	Sequences
crRNA repeat I	5′-**GUUGAAAUCA G**-3′
crRNA repeat II	5′-**GACCAAAAUG G**-3′
crRNA repeat III	5′-**GGGAUUGAAA G**-3′
39-mer crRNA	5′-**AUUGAAAG**CA GGAGGGACCG GAAACACACG GUUGAAGGG-3′
55-mer ssRNA	5′-GUGUGUGUGU AUCAAUCUAU UAAAAUUGUC GUGAAAUGUU-3′
55-mer ssDNA	5′-TCCACCGCCA TAAAGTACG ACGTCCGTCT TCGGTTGTGT GGCTGGAGCT GCTTC-3′
dsRNA (30 bp)	5′-CUCUACGACA UCGGAUCCGA UGUCGUAGAG-3′
dsDNA (112 bp)	*LEE1* promoter region (−60 to +52) in *E. coli* (EPEC)

mutant proteins were purified using the same procedure as was used to purify the wild-type protein. The final purified samples were concentrated to 9 mg ml^{-1}.

2.4. Electrophoretic mobility shift assay (EMSA)

The sequences of the ten different nucleic acids used in this study are tabulated in Table 2. The synthesized RNA strands were purchased from ST Pharm (Republic of Korea). The 5′ ends of RNAs and DNAs were labelled with [γ-^{32}P]-ATP using T4 polynucleotide kinase (Roche). To remove unincorporated [γ-^{32}P]-ATP, the mixture was desalted using an RNase-free Sephadex G-25 column. The labelled probes were then incubated with *Af* Cmr1 for 30 min at 25 or 70°C in a buffer solution consisting of 20 m*M* Tris–HCl pH 7.5, 10 m*M* magnesium acetate, 300 m*M* KCl, 100 ng µl^{-1} BSA, 100 ng µl^{-1} heparin. The total volume of the mixture was 10 µl and the concentration of the probes was 0.1 µ*M*. The mixtures were loaded onto 6%(*w*/*v*) or 15%(*w*/*v*) nondenaturing polyacrylamide gel (40:1) and electrophoresis was conducted at 70 V for 60 min at 25°C in Tris–borate–EDTA buffer (89 m*M* Tris, 89 m*M* boric acid, 2 m*M* EDTA). The results were visualized using a Fuji phosphorimager.

2.5. Nuclease activity assay

Radioactively labelled DNA or RNA probes were mixed with *Af* Cmr1 for 30 min at 25 or 70 °C in the presence or absence of divalent metal ion in a buffer solution consisting of 20 m*M* Tris–HCl pH 7.5, 300 m*M* KCl, 100 ng µl^{-1} bovine serum albumin, 100 ng µl^{-1} heparin. The mixtures were analyzed as described in §2.4.

3. Results and discussion

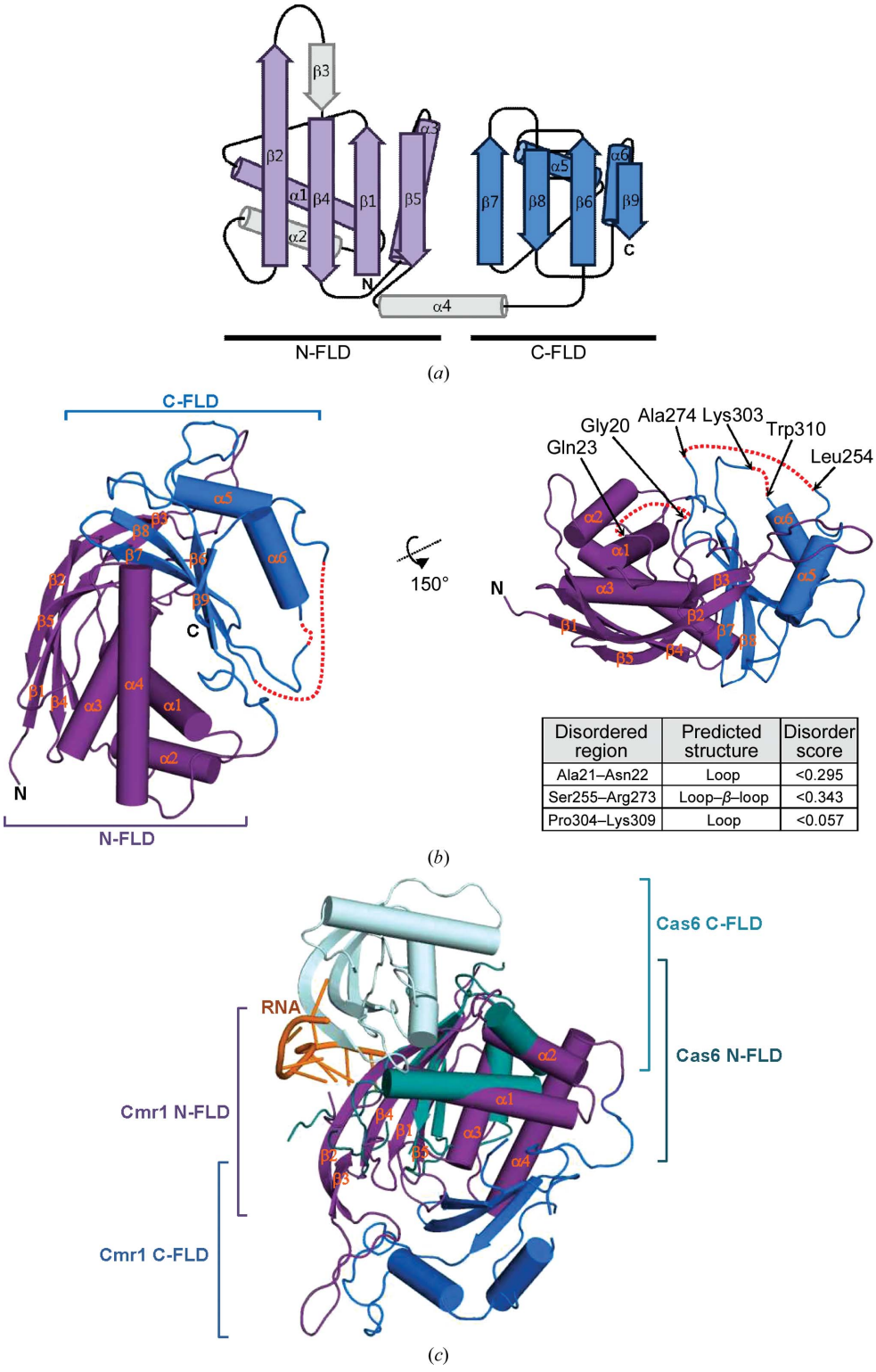
3.1. Structural features of *Af* Cmr1

Recombinant full-length *Af* Cmr1 was produced in *E. coli* and its crystal structure was determined to 2.5 Å resolution (Table 1). The refined *Af* Cmr1 structure is composed of nine β-strands and six α-helices and appears as a single-domain mixed α/β fold with overall dimensions of approximately 48 × 60 × 36 Å. A topology diagram revealed the existence of two $\beta\alpha\beta\beta\alpha\beta$ super-secondary structures, which form a β-sheet with α-helices at one side. This spatial arrangement represents the signature topology of the ferredoxin fold (Fig. 1*a*). In addition to the signature secondary structures, *Af* Cmr1 contains extra α-helices (α2 and α4) and a strand (β3) in the N-terminal domain. α4 spans both the N- and C-terminal domains. Thus, the protein is composed of two ferredoxin-like domains (FLDs) that are tightly associated with each other and barely discernible (Fig. 1*b*). A search for homologous structures with *DALI* (Holm & Sander, 1996) revealed that *Af* Cmr1 is most similar to the structure of Cas6 from *P. furiosus* (PDB entry 3pkm; *Z*-score 7.0), which is a CRISPR RNA-processing enzyme composed of two distinctive FLDs (Wang *et al.*, 2011). The similarity, however, is confined to the N-terminal FLD of *Af* Cmr1, which can be superposed onto the N-terminal FLD of Cas6. Their secondary-structural elements superpose onto each other fairly well, with a root-mean-square deviation (r.m.s.d.) of 2.8 Å for 96 superposed C$^{\alpha}$ atoms (Fig. 1*c*), while the two domains are unrelated in amino-acid sequence. Upon superposition, the C-terminal FLDs of *Af* Cmr1 and Cas6 occupy opposite spatial positions. A *DALI* search with the C-terminal FLD of *Af* Cmr1 revealed that the best match to this domain is the N-terminal domain of ATP phosphoribosyltransferase (PDB entry 2vd3; B. Lohkamp, T. Schweikert & A. J. Lapthorn, unpublished work). While the same folding topology is shared by the two domains, their structural homology is low (*Z* score 4.2), indicating that the functions of the two domains are unrelated.

According to *IUPred* (http://iupred.enzim.hu/), *Af* Cmr1 is not predicted to contain an intrinsically disordered region. However, the *Af* Cmr1 structure contains three disordered segments (residues 20–22, 253–274 and 302–309) whose electron densities are poor or invisible (Figs. 1*b* and 2*a*). Notably, these disordered segments emanate from one face of the protein: the crRNA-binding interface, as described below. It is probable that they undergo disorder-to-order changes upon the binding of other subunit(s) and/or crRNA to this surface.

3.2. Sequence similarity to other Cmr subunits and conserved regions

Using a *BLAST* search (Altschul *et al.*, 1990), a number of proteins annotated as Cmr4, Cmr6 and Csm3 can be retrieved, in addition to many Cmr1 homologues. In particular, the N-terminal FLD of *Af* Cmr1 exhibits sequence homology to the Cmr4 (AF1863) and Cmr6 (AF1861) subunits belonging to the same interference complex. Interestingly, the sequence homology is limited to the C-terminal region of Cmr4 (residues 208–338; 15% identity) and the N-terminal region in the case of Cmr6 (residues 11–199; 21% identity) (data not shown). While the structures of these two subunits are unavailable, these proteins presumably contain at least one FLD. The recently determined structure of the Cmr2–Cmr3 subcomplex derived from *P. furiosus* shows that Cmr3 is composed of the N- and C-terminal FLDs and a middle insertion domain (Shao *et al.*, 2013). Thus, except for the Cmr5 subunit and the hallmark Cmr2 subunit, the other four sub-



Disordered region	Predicted structure	Disorder score
Ala21–Asn22	Loop	<0.295
Ser255–Arg273	Loop–β–loop	<0.343
Pro304–Lys309	Loop	<0.057

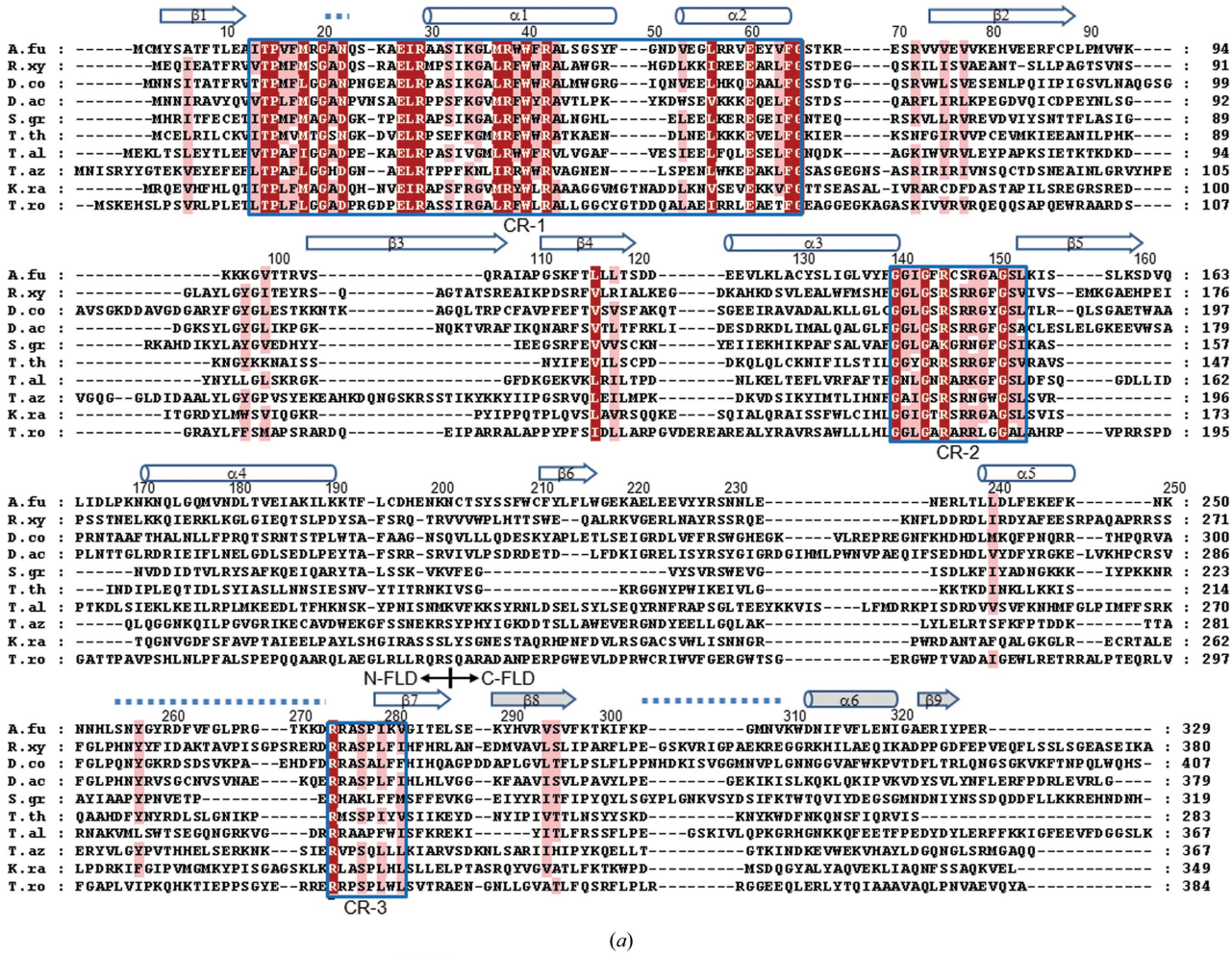
Figure 1
Structural features of *Af* Cmr1. (*a*) Folding topology. *Af* Cmr1 is composed of N- and C-terminal ferredoxin-like domains (N-FLD and C-FLD), each of which contains the signature topology $\beta\alpha\beta\beta\alpha\beta$ shown in violet and blue, respectively. (*b*) Two views of the structure. The secondary-structural elements are numbered in the order of their appearance in the primary sequence. The disordered segments are indicated by dotted lines and the flanking residues are labelled. Their predicted secondary structures and disorder scores are tabulated. (*c*) Structural similarity to Cas6. The structures of *Af* Cmr1 and Cas6 are superimposed. Only the N-terminal FLDs can be superposed. Compared with the N-FLD of Cas6, that of *Af* Cmr1 contains an extra α-helix (α4), which interacts with both the N-terminal and the C-terminal FLDs. The RNA molecule bound to Cas6 is shown in orange.

units Cmr1, Cmr3, Cmr4 and Cmr6 presumably contain at least one FLD.

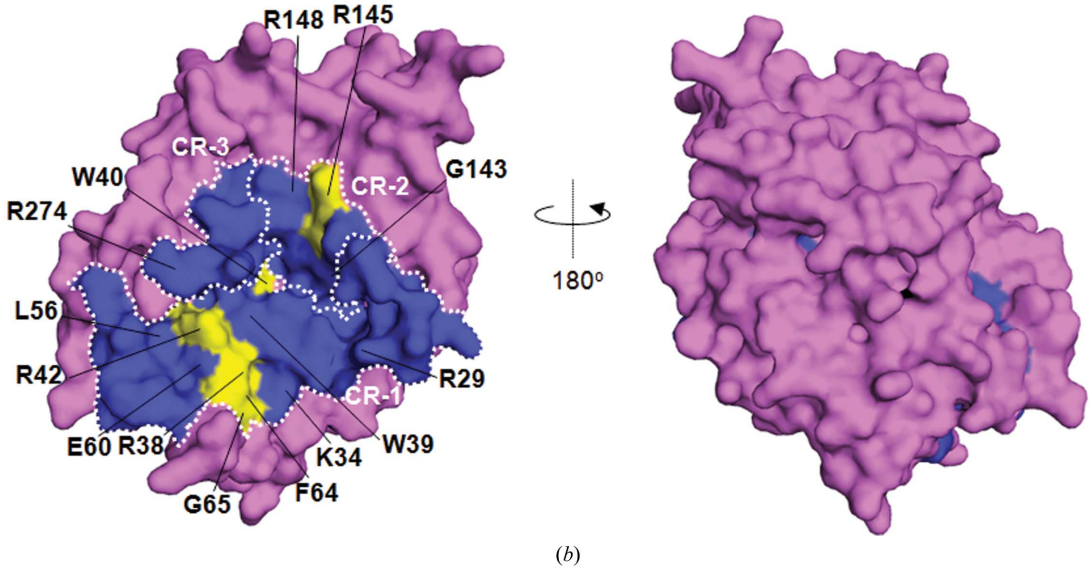
A multiple sequence alignment of the Cmr1 homologues retrieved by *BLAST* revealed that the N-terminal FLD contains two notably conserved regions referred to as CR-1 and CR-2. In contrast, the C-terminal FLD is highly divergent except for having a conserved eight-residue segment, referred to as CR-3 (Fig. 2*a*). Of note, the N-terminal 157-residue segment, which encompasses CR-1 and CR-2, is a conserved domain annotated as Cmr1_III-B in the NCBI Conserved Domain Database. Remarkably, structural mapping of the residues forming the three conserved regions revealed that they are concentrated almost exclusively on one side of the protein (Fig. 2*b*). The surface patch composed by CR-1, CR-2 and CR-3 is rather flat and coincides with the face of the protein from which the three disordered segments extend outwards (Fig. 1*b*). This surface patch contains many invariant or highly conserved basic residues which are exposed to bulk solvent (Fig. 3*a*), suggesting that *Af* Cmr1 might bind nucleic acids through this basic patch.

3.3. Nucleic acid-binding activity of *Af* Cmr1

Postulating that *Af* Cmr1 could interact with a mature crRNA, we performed an EMSA with a 39-mer single-stranded RNA (ssRNA) whose sequence is derived from a repeat and the adjacent spacer sequence of the CRISPR locus 3 in *A. fulgidus* DSM 4304 (Grissa *et al.*, 2007*a*). The assay clearly showed that *Af* Cmr1 binds this crRNA fairly tightly, exhibiting a complete mobility shift of this RNA (Fig. 3*b*). It is noted that the surface patch of *Af* Cmr1 is distinct from the premature



(*a*)



(*b*)

Figure 2
Conserved regions in the *Af* Cmr1 protein. (*a*) Multiple sequence alignment. A total of 55 homologues were retrieved by *BLAST* and aligned using *Clustal Omega* (http://www.ebi.ac.uk/Tools/msa/clustalo/). Of these, ten distantly related sequences were selected and are shown. The conserved regions (CR-1, CR-2 and CR-3) are indicated by blue boxes and the invariant residues are indicated on a red background. The dotted lines indicate disordered regions. A.fu, *Archaeoglobus fulgidus* (gi:11499452); R.xy, *Rubrobacter xylanophilus* (gi: 289192071); D.co, *Diplosphaera colitermitum* (gi:225158929); D.ac, *Desulfotomaculum acetoxidans* (gi:258516135); S.gr, *Saprospira grandis* (gi:424843568); T.th, *Thermotoga thermarum* (gi:338730944); T.al, *Thermocrinis albus* (gi:289547783); T.az, *Treponema azotonutricium* (gi:333996009); K.ra, *Ktedonobacter racemifer* (gi:298243108); T.ro, *Thermomicrobium roseum* (gi:221635522). (*b*) Mapping on the *Af* Cmr1 structure. The three conserved regions (delimited by the white dotted lines) are shown and the surface-exposed invariant or highly conserved residues (>90%) in these regions are labelled and shown in yellow and blue, respectively. Seven of these residues are lysines or arginines.

CRISPR RNA-binding site of Cas6 (Fig. 1*c*), which is a prominent groove between the N- and C-terminal FLDs (Wang *et al.*, 2011). The corresponding region of *Af* Cmr1 is unlikely to be involved in RNA binding because it is a negatively charged surface that lacks a groove-like feature (not shown). To determine whether the basic surface patch of *Af* Cmr1 is the crRNA-binding interface, we produced seven *Af* Cmr1 mutants containing substitutions of highly conserved or invariant arginine residues on this surface with glutamates: R29E, R38E, R42E, R145E, R148E, R274E and R275E. These arginine residues belong to the CR-1, CR-2 or CR-3 regions (Fig. 3*a*). Among the seven mutations, the R29E and R148E mutations exhibited the greatest reduction in the mobility shift of the crRNA (Fig. 3*b*). Fewer, but noticeable, mobility shifts were also observed for the R38E and R42E mutations (Fig. 3*b*). We generated four additional mutants that contained two of the single substitutions: R38E/R145E, R38E/R148E, R42E/R145E and R42E/R148E. Compared with the single mutations R38E, R42E and R145E, the combined mutations R38E/R145E and R42E/R145E caused a noticeably reduced RNA-binding affinity (Fig. 3*b*), indicating an additive effect of the individual mutations. Unlike these mutations of the conserved residues, the mutation of an unconserved positively charged residue, R275E, did not affect the crRNA-binding activity of *Af*

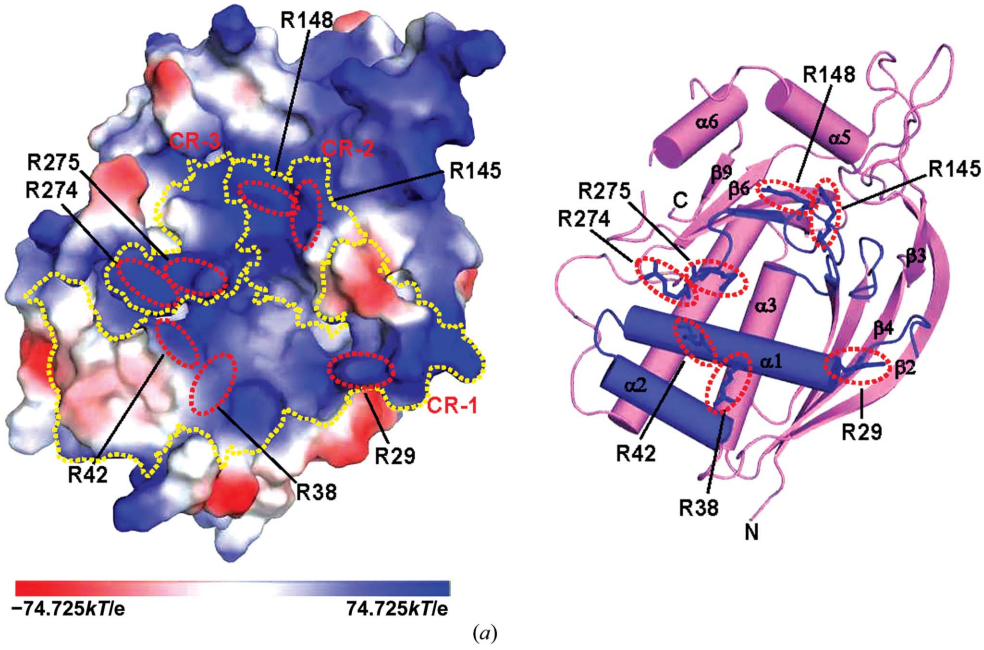


Figure 3
Af Cmr1 binds single-stranded RNA and DNA. (*a*) The presence of a positively charged surface patch. Surface and cartoon models of *Af* Cmr1 are shown in the same orientation as in Fig. 2(*b*). The three conserved regions are indicated by yellow dotted lines in the left panel and in blue in the right panel. The colouring scheme for the electrostatic potentials is shown below the figure. The positions of the arginine residues mutated in this study are indicated. (*b*) *Af* Cmr1 binds crRNA. ^{32}P-labelled 39-mer crRNA (0.3 µ*M*) was incubated with wild-type or the indicated mutant *Af* Cmr1 protein (3 µ*M*) for 30 min at 70°C. The positions of the charge-inversion mutations are shown in (*a*). EMSA was performed on a 15%(*w*/*v*) nondenaturating polyacrylamide gel. (*c*) EMSA with four types of nucleic acids. Increasing amounts of wild-type *Af* Cmr1 (0, 0.4, 0. 75, 1.5 and 3 µ*M*) were incubated with each of the indicated nucleic acids (0.3 µ*M*). EMSA was performed on a 6%(*w*/*v*) nondenaturating polyacrylamide gel.

Cmr1 (Fig. 3*b*). Together, these experiments indicate that *Af* Cmr1 binds crRNA using the conserved positively charged patch that is constituted by the three conserved regions of the protein.

To determine whether *Af* Cmr1 binds other types of nucleic acids, we performed EMSA with 39-mer crRNA, ssDNA, dsRNA and dsDNA with increasing concentrations of the protein. As shown in Fig. 3(*c*), *Af* Cmr1 also binds to ssDNA. In contrast, no apparent binding of the protein to dsRNA or dsDNA was observed. These results indicate that *Af* Cmr1 selectively binds single-stranded nucleic acids. Based on the band intensity, the protein appears to bind ssRNA preferentially over ssDNA (Fig. 3*c*, bottom panels).

3.4. *Af* Cmr1 cleaves single-stranded RNA and DNA

It is as yet unknown which subunit of the Cmr RNA silencing complex possesses the endoribonuclease activity that cleaves target RNA. Based on the structural similarity between the N-terminal FLDs of *Af* Cmr1 and Cas6, we examined whether *Af* Cmr1 has a feature similar to the putative catalytic triad (His46-Tyr31-Lys52) of Cas6 (Wang *et al.*, 2011). *Af* Cmr1 does not contain residues corresponding to the triad residues at spatially similar positions (not shown), nor does it have a bound bivalent metal ion which serves as a cofactor for the catalysis. Whereas a catalytic residue is usually invariant and has a reactive side chain, *Af* Cmr1 does not possess such a residue according to a multiple sequence alignment of the Cmr1 homologues (Fig. 2*a*). Nonetheless, we found that *Af* Cmr1 cleaves 39-mer crRNA in a metal-ion-dependent manner (Fig. 4*a*). At 70°C, this ribonuclease activity was highest and second highest in the presence of Mn^{2+} and Ca^{2+}, respectively. Notably, the effect of Mn^{2+} was not pronounced in the assay performed at 25°C. The ribonuclease activity appeared to be sequence-independent, because *Af* Cmr1 cleaved ssRNAs that were identical to three different *A. fulgidus* CRISPR repeat sequences and also cleaved a 55-mer ssRNA which was unrelated to the crRNA sequences (Fig. 4*b*). *Af* Cmr1 did not cleave dsRNA, which is consistent with its inability to bind dsRNA (Fig. 4*b*). The protein cleaved ssDNA, but less efficiently than it cleaved ssRNA (Fig. 4*b*, right panel, lanes 9 and 10). The observed RNase activity of *Af* Cmr1 is consistent with a recent report suggesting that the cleavage of target RNA might occur near the junction of Cmr1/Cmr6 and Cmr4/Cmr5, which was based on EM images of the *P. furiosus* Cmr holocomplex (Spilman *et al.*, 2013). Since *Af* Cmr1 alone neither bound dsRNA nor cleaved dsRNA under our experimental conditions (Figs. 3*c* and 4*b*), it is enigmatic how Cmr1 in the holocomplex could cleave target RNA that forms dsRNA together with the guide crRNA. Also enigmatic is the location of the active site of the protein. One possibility would be that the main-chain carbonyl O atoms form a metal-binding site and a metal-bound water molecule acts as a catalytic water molecule in the hydrolysis of RNA.

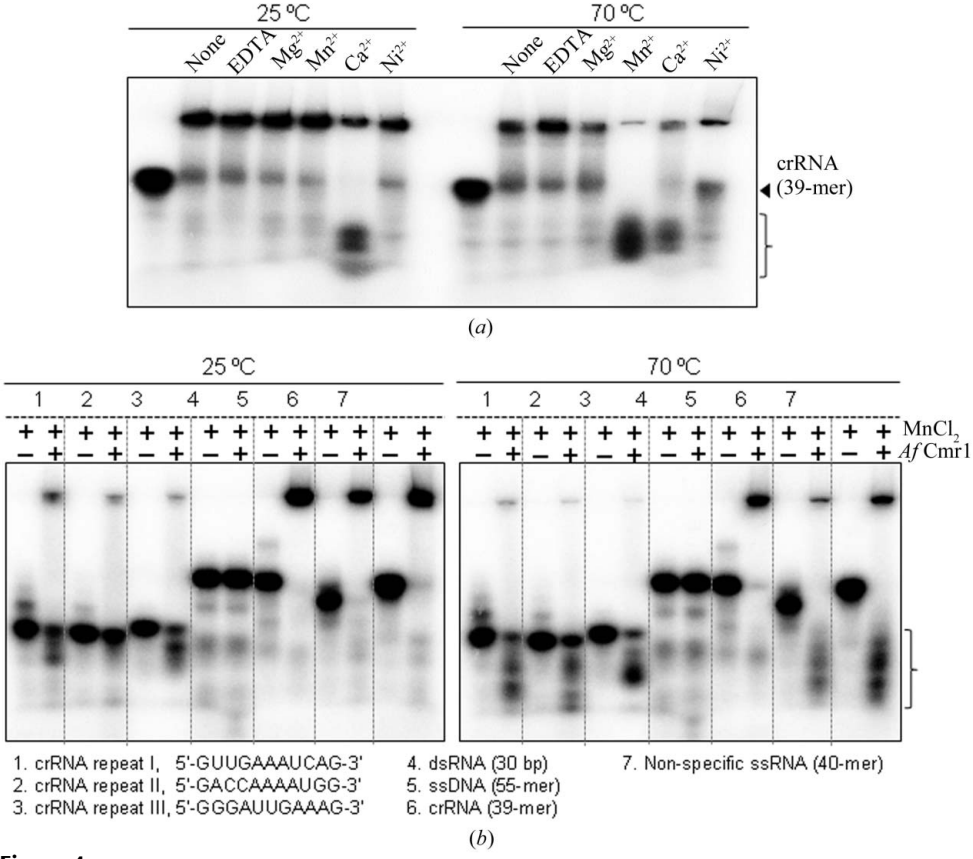


Figure 4
Af Cmr1 exhibits metal-ion-dependent nuclease activity. (*a*) ^{32}P-labelled 39-mer crRNA (0.3 µ*M*) was incubated with *Af* Cmr1 (3 µ*M*) for 30 min at 25 or 70°C in the absence or presence of 10 m*M* $MgCl_2$, $MnCl_2$, $CaCl_2$, $NiCl_2$ or EDTA. (*b*) The indicated RNA or DNA probes (0.3 µ*M*) were incubated with *Af* Cmr1 (3 µ*M*) for 30 min at 25 or 70°C in the presence of 10 m*M* $MnCl_2$. The brackets in (*a*) and (*b*) indicate the degradation products. EMSA was performed on a 15%(*w*/*v*) nondenaturating polyacrylamide gel.

4. Concluding remarks

In this work, we present the first structure of the Cmr1 protein, showing that the protein is composed of two FLDs and that the highly conserved surface patch forms the crRNA-binding interface. Our analyses, together with published structural information, indicate that the predominant structural unit in the Cmr RNA silencing complex is the FLD, which is a module for protein–protein interaction as well as RNA binding. This work also identified Cmr1 as the ribonuclease subunit in the Cmr holocomplex. The precise mechanism of how this complex

slices target RNA is an important question for future investigation.

The authors declare that they have no conflicts of interest with respect to the current study. The X-ray diffraction experiments used beamline 8.3.1 at the Advanced Light Source in the USA and beamline 5C at the Pohang Accelerator Laboratory in the Republic of Korea. This work was supported by grants from the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (2010-0008560) and by the World Class Institute (WCI) Program of the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology of Korea (MEST; NRF grant No. WCI 2009-002) to J-SK, and from the Marine and Extreme Genome Research Center Program funded by the Ministry of Oceans and Fisheries of Korea to B-HO.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.
Aslanidis, C. & de Jong, P. J. (1990). *Nucleic Acids Res.* **18**, 6069–6074.
Babu, M. *et al.* (2011). *Mol. Microbiol.* **79**, 484–502.
Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A. & Horvath, P. (2007). *Science*, **315**, 1709–1712.
Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. (2005). *Microbiology*, **151**, 2551–2561.
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. & van der Oost, J. (2008). *Science*, **321**, 960–964.
Carte, J., Wang, R., Li, H., Terns, R. M. & Terns, M. P. (2008). *Genes Dev.* **22**, 3489–3496.
Cocozaki, A. I., Ramia, N. F., Shao, Y., Hale, C. R., Terns, R. M., Terns, M. P. & Li, H. (2012). *Structure*, **20**, 545–553.
Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J. & Charpentier, E. (2011). *Nature (London)*, **471**, 602–607.
Deveau, H., Garneau, J. E. & Moineau, S. (2010). *Annu. Rev. Microbiol.* **64**, 475–493.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Garneau, J. E., Dupuis, M. È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A. H. & Moineau, S. (2010). *Nature (London)*, **468**, 67–71.
Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. (2012). *Proc. Natl Acad. Sci. USA*, **109**, E2579–E2586.
Gesner, E. M., Schellenberg, M. J., Garside, E. L., George, M. M. & Macmillan, A. M. (2011). *Nature Struct. Mol. Biol.* **18**, 688–692.
Godde, J. S. & Bickerton, A. (2006). *J. Mol. Evol.* **62**, 718–729.
Grissa, I., Vergnaud, G. & Pourcel, C. (2007*a*). *BMC Bioinformatics*, **8**, 172.
Grissa, I., Vergnaud, G. & Pourcel, C. (2007*b*). *Nucleic Acids Res.* **35**, W52–W57.
Hale, C. R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A. M., Glover, C. V. III, Graveley, B. R., Terns, R. M. & Terns, M. P. (2012). *Mol. Cell*, **45**, 292–302.
Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M. & Terns, M. P. (2009). *Cell*, **139**, 945–956.
Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J. A. (2010). *Science*, **329**, 1355–1358.
Holm, L. & Sander, C. (1996). *Science*, **273**, 595–603.
Holton, J. & Alber, T. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 1537–1542.
Jansen, R., Embden, J. D., Gaastra, W. & Schouls, L. M. (2002). *Mol. Microbiol.* **43**, 1565–1575.
Jore, M. M. *et al.* (2011). *Nature Struct. Mol. Biol.* **18**, 529–536.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. (2006). *Biol. Direct*, **1**, 7.
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., van der Oost, J. & Koonin, E. V. (2011). *Nature Rev. Microbiol.* **9**, 467–477.
Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. (2005). *J. Mol. Evol.* **60**, 174–182.
Nam, K. H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisa, M. P. & Ke, A. (2012). *Structure*, **20**, 1574–1584.
Pourcel, C., Salvignol, G. & Vergnaud, G. (2005). *Microbiology*, **151**, 653–663.
Sashital, D. G., Jinek, M. & Doudna, J. A. (2011). *Nature Struct. Mol. Biol.* **18**, 680–687.
Shao, Y., Cocozaki, A. I., Ramia, N. F., Terns, R. M., Terns, M. P. & Li, H. (2013). *Structure*, **21**, 376–384
Spilman, M., Cocozaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R., Terns, M., Li, H. & Stagg, S. (2013). *Mol. Cell*, **52**, 146–152.
Staals, R. H. *et al.* (2013). *Mol. Cell*, **52**, 135–145.
Terwilliger, T. C. (2003). *Methods Enzymol.* **374**, 22–37.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Wang, R., Preamplume, G., Terns, M. P., Terns, R. M. & Li, H. (2011). *Structure*, **19**, 257–264.
Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S. M., Ma, W. & Doudna, J. A. (2009). *Structure*, **17**, 904–912.
Zhang, J., Abadia, E., Refregier, G., Tafaj, S., Boschiroli, M. L., Guillard, B., Andremont, A., Ruimy, R. & Sola, C. (2010). *J. Med. Microbiol.* **59**, 285–294.
Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S. V., Naismith, J. H., Spagnolo, L. & White, M. F. (2012). *Mol. Cell*, **45**, 303–313.
Zhu, X. & Ye, K. (2012). *FEBS Lett.* **586**, 939–945.