Molecular Cell
Article

Structure and Mechanism of the CMR Complex for CRISPR-Mediated Antiviral Immunity

Jing Zhang,[1,5] Christophe Rouillon,[1,5] Melina Kerou,[1] Judith Reeks,[1] Kim Brugger,[2] Shirley Graham,[1] Julia Reimann,[4] Giuseppe Cannone,[3] Huanting Liu,[1] Sonja-Verena Albers,[4] James H. Naismith,[1] Laura Spagnolo,[3,*] and Malcolm F. White[1,*]

[1]Biomedical Sciences Research Complex, University of St Andrews, Fife KY16 9ST, UK
[2]EASIH, University of Cambridge, Addenbrookes Hospital, Cambridge CB2 0QQ, UK
[3]Institute of Structural Molecular Biology and Centre for Science at Extreme Conditions, University of Edinburgh, Edinburgh EH9 3JR, UK
[4]Archaeal Molecular Biology Group, Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse 10, 35043 Marburg, Germany
[5]These authors contributed equally to this work
*Correspondence: laura.spagnolo@ed.ac.uk (L.S.), mfw2@st-andrews.ac.uk (M.F.W.)
DOI 10.1016/j.molcel.2011.12.013

SUMMARY

The prokaryotic *c*lusters of *r*egularly *i*nterspaced *p*alindromic *r*epeats (CRISPR) system utilizes genomically encoded CRISPR RNA (crRNA), derived from invading viruses and incorporated into ribonucleoprotein complexes with CRISPR-associated (CAS) proteins, to target and degrade viral DNA or RNA on subsequent infection. RNA is targeted by the CMR complex. In *Sulfolobus solfataricus*, this complex is composed of seven CAS protein subunits (Cmr1-7) and carries a diverse "payload" of targeting crRNA. The crystal structure of Cmr7 and low-resolution structure of the complex are presented. *S. solfataricus* CMR cleaves RNA targets in an endonucleolytic reaction at UA dinucleotides. This activity is dependent on the 8 nt repeat-derived 5′ sequence in the crRNA, but not on the presence of a protospacer-associated motif (PAM) in the target. Both target and guide RNAs can be cleaved, although a single molecule of guide RNA can support the degradation of multiple targets.

INTRODUCTION

The CRISPR system has recently come to light as a complex mechanism of cell-mediated antiviral immunity (see Horvath and Barrangou, 2010; Karginov and Hannon, 2010; Marraffini and Sontheimer, 2010a for recent reviews). CRISPRs are genomically encoded arrays of short "spacer" sequences (20–72 bp, depending on the species), each flanked by a repeat sequence with an average length around 25–30 nt. CRISPR arrays are transcribed by the cellular RNA polymerase and processed to generate small crRNAs by nucleolytic cleavage within the repeat sequences (Brouns et al., 2008; Carte et al., 2008; Deltcheva et al., 2011). Processed crRNAs are utilized by a variety of CRISPR-associated (CAS) protein complexes as a guide RNA to target (and degrade the nucleic acid of) invading genetic entities with complementary nucleic acid sequences. This defensive "interference" process works in tandem with an adaptive "capture" process that allows the incorporation of new spacer sequences derived from viruses into the genomic CRISPR arrays.

The viral DNA sequences that become incorporated into CRISPR arrays are known as "protospacers" (Horvath et al., 2008). Protospacers are derived from both coding and noncoding regions of viral genomes, suggesting that the viral DNA, rather than RNA, is targeted by the process that captures new spacers (Horvath et al., 2008; Shah et al., 2009). Examination of the sequence context of protospacers revealed the presence of a conserved "protospacer-associated motif" (PAM) consisting of a di- or trinucleotide signature, immediately adjacent to the protospacer sequence (Bolotin et al., 2005; Deveau et al., 2008; Horvath et al., 2008; Mojica et al., 2009). The presence of a PAM is important for the recognition and restriction of invading mobile DNA elements (Deveau et al., 2008; Gudbergsdottir et al., 2011).

CAS protein complexes have recently been classified into three main subtypes (Makarova et al., 2011b). In *Escherichia coli*, the type I-E complex, commonly known as "CASCADE," consists of five protein subunits (CasA-E). CASCADE processes CRISPR transcripts into ~57 nt crRNAs and uses them to recognize invading viral DNA, which is subsequently cleaved by Cas3 (Brouns et al., 2008). A similar complex (type I-A) with several conserved subunits has been described in *Sulfolobus solfataricus* (Lintner et al., 2011). Many archaea and some bacteria also encode a type III-B system, known as the CMR complex (Haft et al., 2005). In the euryarchaeon *Pyrococcus furiosus*, the Cmr1-6 proteins have been purified as a complex that uses crRNA to target RNA (presumably viral mRNA in vivo), cleaving it with a molecular ruler mechanism guided by the 3′ end of the crRNA (Hale et al., 2009).

Here, we report the purification and characterization of the CMR complex from *S. solfataricus*. There are seven subunits, comprising Cmr1-7 and a crRNA component. Deep sequencing reveals a biased composition for the crRNA, which is largely derived from 2 of the 6 CRISPR loci. The crystal structure of the Cmr7 subunit has been solved and consists of a protein fold with a conserved surface that may mediate molecular

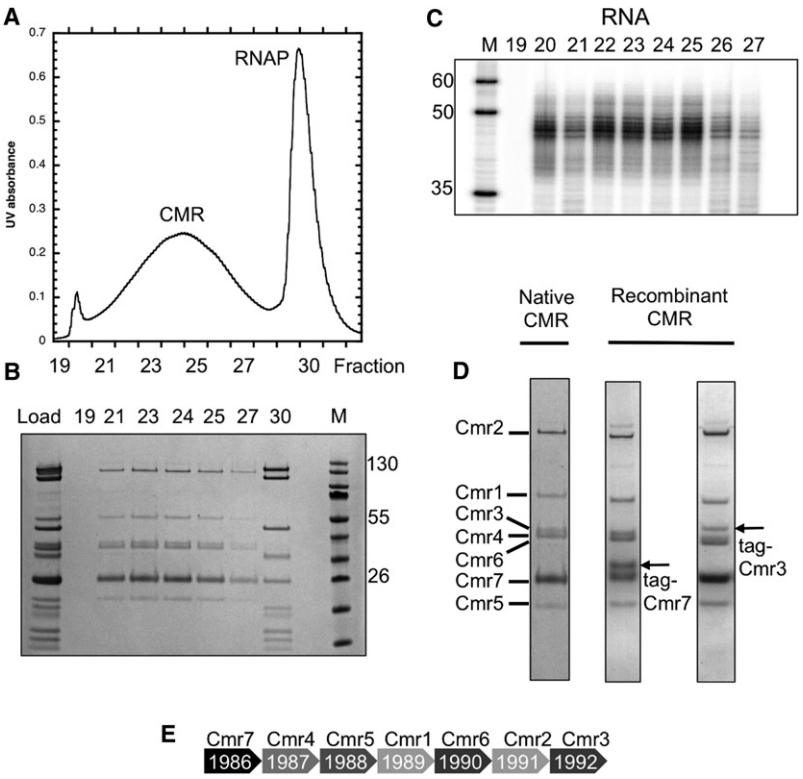


Figure 1. Purification of the CMR Complex of *S. solfataricus*

(A) UV trace showing fractions eluting from final MonoQ column, with CMR and RNAP complexes resolved.

(B) SDS-PAGE analysis of fractions from MonoQ column, showing separation of RNAP (fraction 30) from the 7-subunit CMR complex.

(C) Denaturing gel electrophoresis of end-labeled nucleic acid reveals the presence of RNA copurifying with the CMR complex. The size range centered on 46 nt corresponds to a spacer with an 8 nt repeat-derived 5′ tag.

(D) Comparison of native and tagged versions of the CMR complex purified from *S. solfataricus*. Both tagged and untagged versions of Cmr7 are apparent, reflecting its higher stoichiometry in the complex.

(E) Mapping of Cmr1–7 onto the gene locus *sso1986* to *sso1992*.

interactions. The EM structure of the full CMR complex and a defined subcomplex are presented. We demonstrate that *S. solfataricus* CMR (SsoCMR) utilizes a sequence-dependent RNA cleavage mechanism without a molecular ruler.

RESULTS

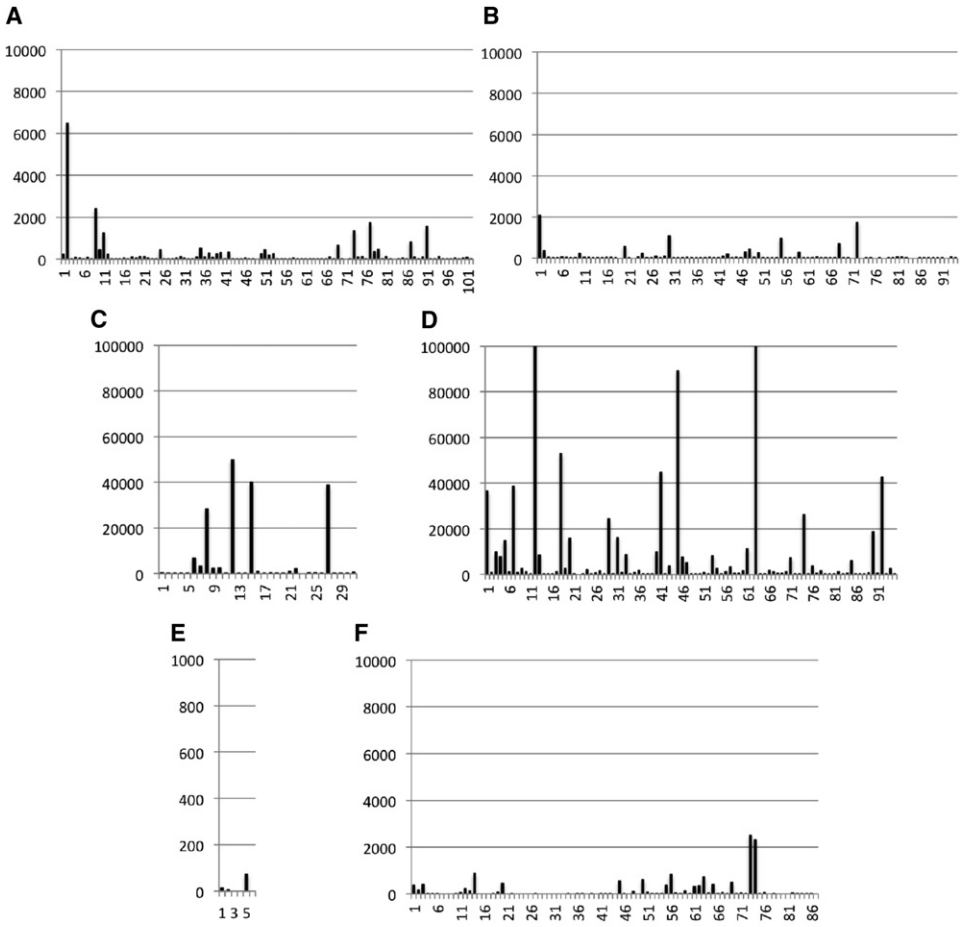
Purification of the CMR Complex from *S. solfataricus*

The native CMR complex was purified from *S. solfataricus* using four sequential column chromatography steps (heparin, gel filtration, MonoS, and MonoQ). At each stage, column fractions were checked for the presence of the Cmr7 subunit by "dot blot" with a specific polyclonal antibody. SsoCMR copurified with the cellular RNA polymerase through the first three columns and was separated by the final anion exchange step (Figure 1A). The purified complex contained seven subunits corresponding to the products of genes *sso*1986 through *sso*1992. Subunits 1–6 were judged present at a 1:1 stoichiometry; densitometric analysis suggested that the Cmr7 subunit was present at a stoichiometry of three dimers in each complex (Figures 1 and S1). This was consistent with an overall size for the complex of 415 kDa (or 430 kDa including the RNA component), explaining the copurification on gel filtration with the 410 kDa RNA polymerase. The presence of RNA, with a variable fragment size centered on 46 nt, was confirmed (Figure 1C). This was in good agreement with the size of the crRNA species isolated from *P. furiosus* CMR (PfuCMR) (Hale et al., 2009), consistent with the presence of a spacer sequence of variable length with a CRISPR repeat-derived 8 nt tag at the 5′ end. By expressing subunits (Cmr7 or Cmr3) of SsoCMR with a polyhistidine tag in *S. solfataricus*, the intact complex could be isolated by incorporating an affinity chromatography step (Figure 1D).

RNA Content of the CMR Complex

To determine the specific characteristics of the RNA component of SsoCMR, isolated RNA was cloned and deep-sequenced. We mapped 1.88 million sequence reads of 36 nt length onto the *S. solfataricus* spacers (Table S1). Analysis of the start and stop positions of the sequence reads for a subset of highly represented spacers from the A and D loci revealed the presence of sequence corresponding to the 5′ tag derived from the CRISPR repeat, with a clear demarcation at the eighth nucleotide, which corresponds to the site of cleavage by Cas6 (Carte et al., 2008; Lintner et al., 2011) (Figure S2). The 3′ ends of the sequenced RNA were more variable. Some spacers such as A2 and D43 displayed a short 3′ handle, while others appeared to have very little repeat-derived sequence at the 3′ end. Overall, this fits the suggestion that crRNA is processed by the Cas6 endoribonuclease followed by exonucleolytic digestion of the 3′ end (Hale et al., 2009). By contrast, crRNA isolated from the *S. solfataricus* CASCADE complex still includes the 3′ repeat-derived sequence (Lintner et al., 2011), suggesting that crRNAs are differentially processed depending on their ultimate destination in CASCADE or CMR.

S. solfataricus P2 has six CRISPR loci, denoted A–F, ranging from 8 to 103 repeats in length (Lillestøl et al., 2006). Whole-genome transcription data suggested that loci A, B, C, and D are all highly transcribed, while E and F are very weakly transcribed (Wurtzel et al., 2010). Examples of spacers derived from all six loci were detected in the CMR complex, but the distribution was highly biased. The majority of CMR-bound crRNAs were derived from locus D, followed by locus C, with significant underrepresentation from the other four loci (Figure 2). Within a locus, spacer representation in the CMR complex was highly variable, with numbers of sequenced crRNAs from adjacent spacers frequently differing by several orders of



CRISPR locus	No. of spacers	Repeat	% of total spacers	% of crRNAs in CMR	Ratio observed/ expected
A	102	GATTAATCCCAAAAGGAATTGAAAG	24	1.3	0.05
B	94	GATTAATCCCAAAAGGAATTGAAAG	23	0.6	0.03
C	31	GATAATCTCTTATAGAATTGAAAG	7.4	9.6	1.3
D	95	GATAATCTCTTATAGAATTGAAAG	23	88	3.8
E	7	GATAATCTACTATAGAATTGAAAG	1.7	0.005	0.003
F	88	GCTAATCTACTATAGAATTGAAAG	21	0.7	0.03

Figure 2. Distribution of crRNA Bound by the *S. solfataricus* CMR Complex

(A–F) Examples of crRNAs from all six CRISPR loci were observed, with a clear bias toward locus D, followed by locus C. The individual plots for each locus show that crRNA representation was highly variable, with adjacent spacers represented at levels that often varied by several orders of magnitude. For each graph, the x axis represents the position of each spacer in the locus and the y axis represents the number of sequenced matches to each spacer. 88% of the total sequence reads were derived from locus D, which represents a nearly 4-fold overrepresentation compared to the proportion of crRNAs encoded by that locus. CRISPR loci E and F, which are poorly transcribed, were significantly underrepresented, as expected. However, crRNAs from CRISPRs A and B, which are highly transcribed and thought to be actively adding spacers in vivo, are also significantly underrepresented in the CMR complex. The table shows the properties and representation in CMR of each CRISPR locus.

magnitude (Figure 2). In particular, one spacer (D63) accounted for 45% of the sequence reads, and the top ten spacers accounted for 79% of the total reads mapped to spacers (Table S2). Reverse transcripts of CRISPR arrays have been detected in *Sulfolobus* species (Lillestøl et al., 2009), but their significance remains unclear. Only 209 RNAs corresponding to reverse transcripts were sequenced, corresponding to 0.01% of the total (Table S1).

The Crystal Structure of Cmr7

The crystal structure of Cmr7 (Sso1986) was solved to 2.05 Å resolution (PDB 2X5Q) (Figure 3). Sso1986 exhibits a fold consisting of six β sheets and four α helices, forming a dimer with a buried surface area of 1177 $Å^2$ and a concave face (Figures 3 and S3A). Homologs of Cmr7 are only detectable in the *Sulfolobales*, but given the rapid evolution of the CRISPR system, it is possible that distant homologs sharing this fold exist

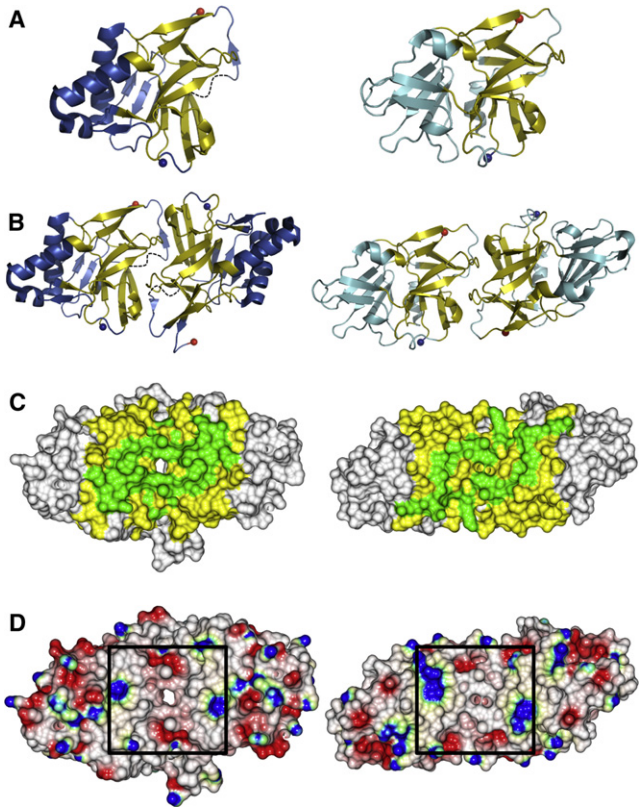


Figure 3. Crystal Structures of Two Members of the Cmr7 Family, Viewed from the Concave Face of the Dimer

(A) The Cmr7 proteins Sso1986 (left) and Sso1725 (right) contain a structurally conserved core (yellow) and a variable region (blue and cyan for Sso1986 and Sso1725, respectively). The β13-β14 loop of Sso1986 is disordered and is represented as a dashed, black line. The N and C termini are represented as blue and red spheres, respectively.

(B) Sso1986 and Sso1725 both form dimers, and the structurally conserved core is located at the dimer interface. The interface itself is also conserved between the two proteins.

(C) The structurally conserved residues (green) and secondary structure (yellow) are located close to the dimer interface with a significant proportion positioned at the concave face.

(D) Electrostatic surface images show that the regions of the concave face proximal to the dimer interface in both proteins (black box) have broad similarities.

Table 1. Data Collection and Refinement Statistics for Sso1725

Data collection	Native	Anomalous
Wavelength (Å)	0.97	1.60
Space group	$P2_12_12_1$	$P2_12_12_1$
a, *b*, *c* (Å)	77.75, 90.29, 111.65	77.71, 91.08, 111.94
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	47.48–2.08 (2.13–2.08)	77.11–2.54 (2.61–2.54)
I/σI	19.2 (2.2)	20.7 (2.8)
R_{merge}	0.04 (0.57)	0.06 (0.71)
Completeness	97.8 (86.1)	99.9 (99.8)
Multiplicity	4.7 (3.7)	7.8 (7.9)
Anomalous completeness	-	99.9 (99.7)
Anomalous multiplicity	-	4.1 (4.1)
Refinement	Sso1725	
R_{work}/R_{free}	0.199/0.231	
Mean B value (Å^2)		
All atoms	59.442	
Protein	59.481	
Water	54.758	
Root-mean-square deviations		
Bond lengths (Å)	0.01	
Angles (°)	1.31	

Each data set was collected on a single crystal at 100 K. Statistics are presented as averages with values for the highest-resolution shell in parentheses. R_{free} was calculated from a random 5% of the reflection data that was omitted from the subsequent refinement.

more widely. To help determine the most important features of Cmr7, we solved the structure of a second Cmr7 subunit, Sso1725. The sequence similarity between the two proteins is low (19% identity), and the crystal structure of Sso1725 (solved to 2.08 Å resolution; Table 1) shows limited structural similarity (Cα rmsd of 2.39 Å over 121 atoms for each monomer). The majority of the residues conserved in the six known orthologs of Cmr7 are located in the dimer interface and concave face (Figures 3C and S3). The electrostatic surfaces of the two proteins show broad similarities at the dimer interface, with symmetrical patches of negative charge at the poles and positive charge at the edges (Figure 3D). Given this level of conservation on only one face of the protein, we hypothesize that this region is important for function, possibly as a binding site for other CMR subunits. The stoichiometry of three dimers of Cmr7 per CMR complex is most easily accommodated if they form a trimeric (pseudo-hexamer) structure in the context of the complex.

Electron Microscopy Reveals the Architecture of the CMR Complex

To elucidate the three-dimensional architecture of the CMR complex, we performed electron microscopy coupled to single-particle analysis experiments for the full complex and additionally for Cmr2/Cmr3/Cmr7 subcomplex, devoid of RNA (Figures 4 and S4). Projections of the maps matched with 2D averages assigned the same Euler angles (Figure S4), highlighting the validity of the maps. The resolution for both maps was ~25 Å, calculated at 0.5 Fourier Shell Correlation (3 sigma). The CMR complex exhibited cavities, compatible with an RNA threading machine. There was no obvious similarity to the "seahorse" structure of *E. coli* CASCADE (Jore et al., 2011). The Cmr2/Cmr3/Cmr7 subcomplex, which lacked bound crRNA, had overall dimensions of 90 × 120 × 110 Å, organized in a clamp or "crab claw" structure (Figure 4A). The intact CMR complex with loaded crRNA had overall dimensions of 160 × 120 × 110 Å with an upper "crab claw" connected to a protruding region (Figure 4B). These dimensions were compatible with the expected molecular masses of ~290 and ~430 kDa, respectively. The Cmr2/Cmr3/Cmr7 subcomplex fitted well to the upper region of intact CMR (Figure 4C), consistent with a role as

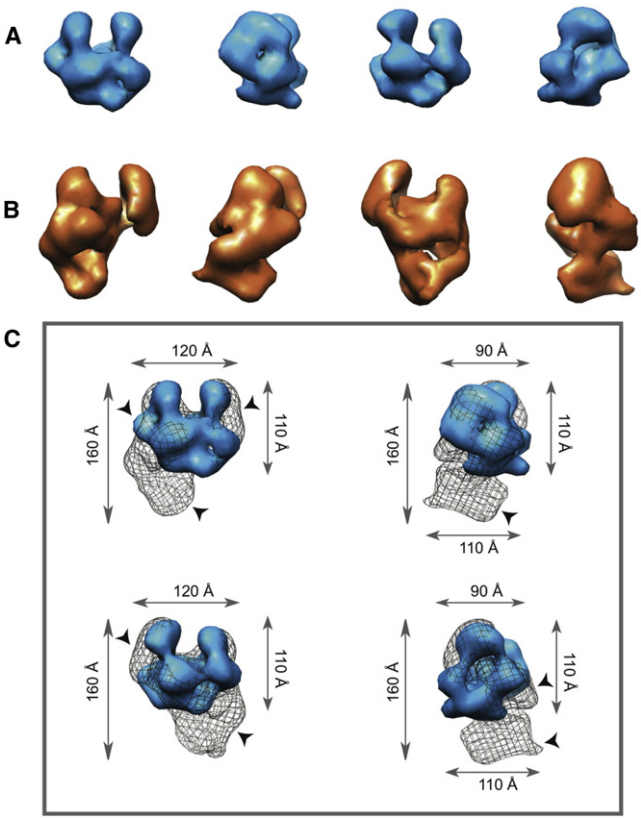


Figure 4. 3D EM Visualization of CMR Complex

(A) Surface representation of the Cmr2/Cmr3/Cmr7 subcomplex devoid of crRNA.

(B) Surface representation of the full CMR complex with bound crRNA.

(C) Superposition of Cmr2/Cmr3/Cmr7 (blue surface) on CMR/RNA (black mesh). Black arrowheads point to regions of additional density on the full CMR complex with bound crRNA compared to Cmr2/Cmr3/Cmr7. Gray arrows indicate dimensions in angstroms.

a scaffold for the assembly of the other Cmr subunits around the periphery (black arrows in Figure 4C).

Ribonuclease Activity of SsoCMR

We tested the ability of SsoCMR to recognize and cleave RNA targets corresponding to spacers A1 and D63 in vitro. Both RNA targets were cleaved efficiently when a cognate crRNA (guide RNA) with an 8 nt 5′ tag was present (Figure 5). Manganese ions were essential for this activity, and magnesium could not substitute. ATP was not essential, but clearly stimulated the cleavage reaction (Figure S5). No crRNA-directed cleavage of DNA targets was observed (data not shown). To rule out the possibility of activity from a contaminating ribonuclease, the CMR complex was immunodepleted using antibodies raised against the Cmr7 subunit. Immunodepletion abolished the nuclease activity, suggesting strongly that the activity was associated with the CMR complex (Figure S5A). As a further control, native untagged SsoCMR purified from *S. solfataricus* cell extract by immunoprecipitation using the anti-Cmr7 antibody had the same activity as the column-purified, tagged protein complex (Figure S5B).

Features of Guide and Target RNAs Important for Cleavage by CMR

The sequence and structural requirements of RNA cleavage by CMR were investigated by constructing a range of target and guide RNA molecules based on the spacer A1 and D63 sequences. To test for a molecular ruler mechanism as observed for PfuCMR, we reduced the length of the guide RNA (Figure 5A, panel ii). The cleavage sites did not move in register with the 3′ end of the guide RNA, suggesting that *Sulfolobus* and *Pyrococcus* CMR differ fundamentally in this respect. Deletion of the repeat-derived 5′ tag from the guide RNA abolished cleavage activity (Figure 5A, panel iii) and could not be rescued by substitution with a 5′ 8A sequence (panel iv), showing that this 5′ tag sequence is essential for cleavage, ruling out the possibility of a contaminating nuclease. The presence of an unpaired flap at the 3′ end of the target RNA was also required for activity (panel vi). This corresponds to the position of the PAM that is essential for cleavage of viral DNA targets by CASCADE (Gudbergsdottir et al., 2011; Lintner et al., 2011). However, for target RNA cleavage by CMR, the PAM sequence at this position was not essential, as a 6A sequence could substitute (Figure 5B, panel vii).

In addition to the target RNA, the guide RNA strand could also be cleaved by SsoCMR (Figure 5C, panel viii). Cleavage of the guide RNA was dependent on the presence of a 3′ overhang on the target RNA (Figure 5C, panel ix, xi). Guide and target were cleaved at approximately equal rates when present at an equimolar ratio, but at ratios of 20:1 or 5:1 excess of target over guide, the guide RNA was cleaved significantly more slowly. Under these conditions, multiple turnover cleavage of the target RNA was observed, suggesting that cleavage of the guide RNA was not essential for catalysis (Figure S5D).

Sequence-Specific Cleavage by SsoCMR

The cleavage patterns observed for SsoCMR suggested a sequence- or structure-specific component to the activity. Sequence mapping suggested that strong cleavage always occurred at a UA dinucleotide in both the A1 and D63 target RNAs and the A1 guide RNA (Figure 5). Weaker cleavage was observed at UU dinucleotides. RNA cleavage by SsoCMR resulted in products with 3′-hydroxyl termini that could be extended by PolyA polymerase (Figure S6A). This is similar to the metal-dependent RNaseH-type activity observed for Piwi and Argonaute (Hutvagner and Simard, 2008). In contrast, the (metal-independent) Cas6 endonuclease yields 3′-cyclic phosphate products that are not extended by PolyA polymerase (Figure S6B). PfuCMR is also reported to generate 3′-cyclic phosphate products (Hale et al., 2009), another distinction between the two enzymes. To map the cleavage site of SsoCMR precisely, the cleavage at site 2 in the D63 target RNA was compared to a synthetic oligonucleotide terminating after the relevant UA dinucleotide (Figure S6C). The cleavage product generated by SsoCMR was 1 nt shorter than the oligonucleotide, consistent with cleavage at this position (and by extension at the other sites) as occurring at the center of the UA sequence.

To examine the importance of sequence for CMR-mediated cleavage of RNA, a D63-derived target RNA with only one UA site, corresponding to position 2, was synthesized. In the

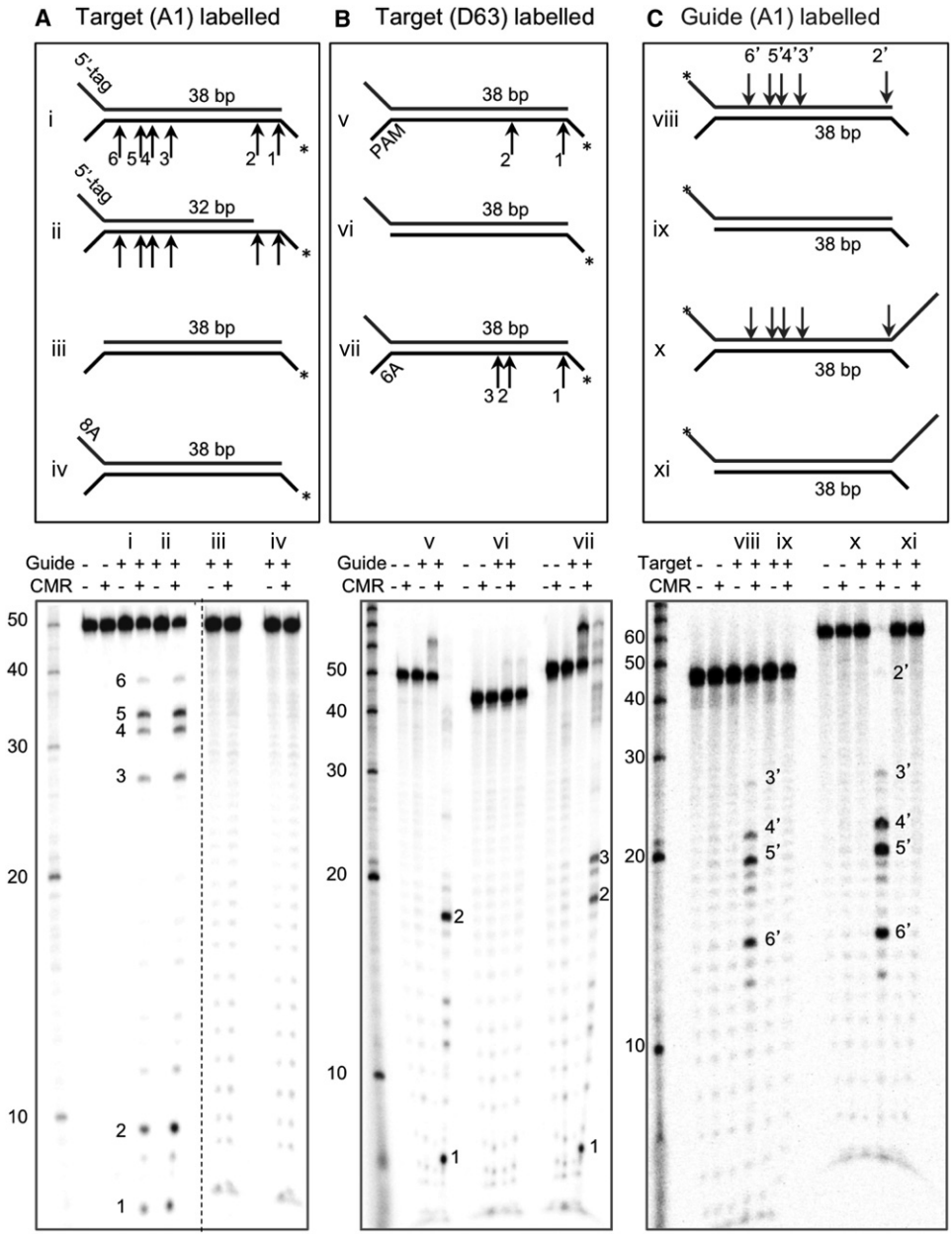


Figure 5. Characterization of the *S. solfataricus* CMR Complex Activity In Vitro

(A) Radiolabeled target RNA (5 μM) corresponding to spacer A1 was incubated with the Cmr7-tagged SsoCMR complex (0.5 μM) and guide RNA (1 μM). Cleavage of the target RNA was observed at six sites (labeled 1–6) (i). 3′ truncation of the guide RNA did not change the cleavage pattern of the target RNA, ruling out a molecular ruler mechanism (ii). The 8 nt 5′ tag was essential for cleavage activity, as either deletion (iii) or replacement with an 8A sequence (iv) abolished nuclease activity.

(B) The D63 target (0.5 μM) was cleaved in the presence of Cmr7-tagged SsoCMR complex (0.5 μM), cognate guide RNA (0.1 μM) (v). Deletion of the unpaired 3′ end of the target RNA, which corresponds to the position of the PAM sequence, abolished activity (vi). The presence of an unpaired 6A sequence at this position restored activity (vii).

(C) Radiolabeled guide RNA corresponding to spacer A1 (3 μM) was incubated with the Cmr7-tagged SsoCMR complex (0.5 μM) and cognate target RNA (1 μM). Cleavage was observed at up to five positions, labeled 2′ to 6′ (viii). This activity was dependent on the presence of the 3′ unpaired spacer sequence (ix) and was not influenced by the presence of an additional unpaired extension at the 3′ end of the guide RNA (x, xi). For each figure part, the 5′ end-labeled RNA strand is indicated with an asterisk. Labeled decade RNA markers (Ambion) are shown.

presence of the CMR complex and a cognate guide RNA sequence, this target was cleaved strongly at position 2, with only weak background cleavage elsewhere (Figure 6B). The same reaction product was observed when pairing this target with the wild-type guide, despite the presence of mismatches at the mutated UA sites. A modified version of the D63 guide

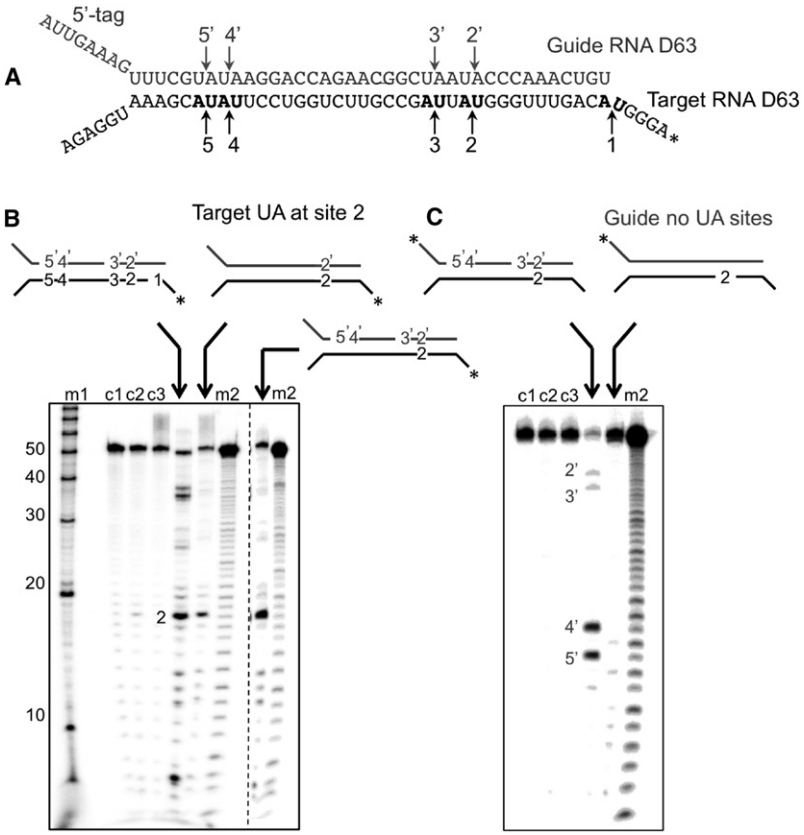


Figure 6. The *S. solfataricus* CMR Complex Cleaves RNA Selectively at UA Sites

(A) Sequence map of the D63 target and guide RNA.

(B) A D63 target oligonucleotide ($D63_{1UA}$, 0.5 μM) with UA cleavage sites 1, 3, 4, and 5 mutated to UG, was labeled and incubated with the Cmr3-tagged CMR complex (0.5 μM) and a cognate guide RNA ($crD63_{1UA}$, 0.05 μM). Cleavage of the target RNA was only observed at the single remaining UA site, site 2. The same target paired with the wild-type guide RNA gave identical cleavage products despite the presence of three mismatches.

(C) A D63 guide RNA sequence with all four UA sites mutated to CA ($crD63_{0UA}$, 0.5 μM) was labeled and incubated with the Cmr3-tagged CMR complex (0.5 μM) and a target RNA containing a single UA site at position 2 ($D63_{1UA}$, 0.05 μM). No significant cleavage of this target sequence was observed. The standard guide RNA (crD63) was cleaved at all four UA sites under the same conditions, despite the presence of mismatches at three positions in the RNA duplex. Control lanes for all gels are: m1, Ambion Decade markers; c1, labeled RNA alone; c2, labeled RNA and CMR; c3, both RNA strands without CMR. Asterisks indicate the 5′ RNA end labeled by ^{32}P.

RNA lacking any UA sites was not cleaved by CMR in the presence of a target RNA with 1 UA site (Figure 6C). The presence of mismatches in the guide RNA opposite the target cleavage site did not abolish cleavage (Figure S6D), consistent with the mismatch tolerance observed previously for CASCADE (Gudbergsdottir et al., 2011; Semenova et al., 2011). Overall, these data confirmed that sequence-dependent cleavage of both the guide and target RNAs is a defining feature of SsoCMR and demonstrated that cleavage of the guide is not necessary for cleavage of the target, consistent with the observation of multiple turnover by the complex.

DISCUSSION

crRNA Content of SsoCMR

There are six CRISPR loci in *S. solfataricus* P2, named A–F (Lillestøl et al., 2006). Deep sequencing of 1.88 million crRNAs isolated from the CMR complex revealed a highly biased distribution, with 98% of the total crRNAs derived from locus D and C, an overrepresentation of 3.8- and 1.3-fold, respectively. The underrepresentation of crRNAs from highly transcribed loci A and B may be explained by differences in the processing of repeats by Cas6. It is notable that the CRISPR repeat sequence associated with the A and B loci is longer than those of C–F, which could provide a plausible basis for differential processing by Cas6. As *S. solfataricus* encodes five different Cas6 proteins, these may be specialized for the cleavage of particular classes of CRISPR repeat and/or may interact differently with the multiple CASCADE and CMR complexes in this organism. Another possibility is that differences in the removal of the repeat-derived sequence 3′ of the spacer by an unknown nuclease after Cas6 processing (Hale et al., 2009) influence incorporation efficiency.

The extreme variation in crRNA sequence numbers obtained from adjacent spacers in all the CRISPR arrays was also unexpected. In some cases, this is probably explained by the presence of internal promoter sequences encoded by adjacent spacers, driving higher transcript levels for particular regions of crRNA. Another variable that could influence crRNA processing is the potential for the formation of stable folded RNA structures following transcription that could influence crRNA processing by Cas6 and loading into the CMR complex. Each spacer has a unique sequence that confers a particular capacity to fold on the local crRNA structure. To address this further, we analyzed the thermodynamic stability of folded RNA structures for each crRNA derived from CRISPR locus C and plotted these against the frequency of occurrence of each crRNA in the CMR sequencing sample (Figure S2). The crRNAs with the potential to fold into the most stable structures were clearly not highly represented in the data set, while the most highly represented sequences had similar, modest folding propensity. This is suggestive of an influence of the secondary structure of the crRNA locally on the efficiency of cleavage by Cas6 and possibly on the loading of crRNA into the CMR complex. Such effects may be ameliorated in the thermophiles by the high growth temperature, but could constitute a significant problem in temperature mesophiles. In organisms such as *E. coli*, the repeat sequence is palindromic, folding into a stable hairpin structure that is recognized by the CasE nuclease (Brouns et al., 2008). This may be an evolutionary mechanism that helps impose an

ordered secondary structure on the crRNA to avoid problems due to folding of spacer sequences. In contrast, Cas6, which is present in many thermophiles, binds unstructured crRNA repeats (Wang et al., 2011).

Structure of the CMR Complex

The EM envelope of the Cmr2/Cmr3/Cmr7 subcomplex bears some similarity to the clamp region of RNA polymerase. In particular, the distance between the two sides of the "claw" is ~30 Å wide, and their length is ~40 Å. It is tempting to view this feature as a dsRNA binding cleft, particularly as Cmr2 is assumed to harbor the active site of the CMR complex. In support of this, the cleft is somewhat deeper in the Cmr2/Cmr3/Cmr7 subcomplex, which contains no bound crRNA, in comparison to the full complex. The lack of crRNA in the Cmr2/Cmr3/Cmr7 subcomplex fits with a presumed role for the RAMP-containing Cmr subunits (Cmr1, Cmr4, and Cmr6) in RNA binding (Makarova et al., 2011a). This is consistent with the recent prediction that Cmr3 is most closely related to the Cas5 subunit of CASCADE, while Cmr1, Cmr4, and Cmr6 are closer matches to the Cas7 subunit (CasC in *E. coli*), which is known to bind crRNA (Makarova et al., 2011a). Compared to Cmr2/Cmr3/Cmr7, the additional Cmr subunits are distributed mainly at the front and at the bottom of the complete CMR complex and may form a structure related to the crRNA binding CasC backbone of CASCADE. Visualization of the path of RNA in the CMR structure remains a key aim to help elucidate the mechanism and molecular organization.

The Mechanism of SsoCMR in Viral RNA Cleavage

We have demonstrated that the SsoCMR complex cleaves target RNA in a sequence-specific manner that is dependent on the presence of a guide crRNA with a cognate sequence and an 8 nt repeat-derived tag at the 5′ end. Cleavage occurs at UA dinucleotides, generating products with 3′-OH and 5′-phosphate ends, like those produced by Argonaute and Piwi (Hutvagner and Simard, 2008). The cleavage sequence is extremely common in the *Sulfolobales* and their viruses. Of the >400 spacers in the CRISPR loci of the *S. solfataricus* P2 genome, only 11 lack a UA dinucleotide. The palindromic nature of this dinucleotide ensures that an equivalent UA sequence is always present in the guide RNA. Cleavage of both the target and guide RNA at UA sites was observed, with the guide cleaved significantly more slowly at high ratios of target:substrate. Under these conditions, one guide RNA molecule could support the cleavage of several target RNAs, demonstrating that cleavage of the guide is not essential for target RNA destruction.

Although not yet shown directly, the N-terminal permuted HD nuclease domain present in Cmr2 is generally assumed to be the nuclease site of the CMR complex. The distantly related HD domain present in Cas3 has been shown capable of cleaving both RNA and DNA, although they are specific for single-stranded nucleic acids and generate products with 3′-cyclic phosphates (Mulepati and Bailey, 2011; Beloglazova et al., 2011). It is also possible that the RNA cleavage activity resides elsewhere in the Cmr2 subunit or in one or more of the RAMP-containing subunits (Cmr1, Cmr2, Cmr4, and Cmr6), which are distantly related to Cas5, Cas6, and Cas7 (Makarova et al., 2011a).

We observed no requirement for a PAM sequence adjacent to the protospacer in RNA targets. This is a marked difference from the DNA targeting CASCADE in archaea, which only cleaves substrates containing a PAM (Gudbergsdottir et al., 2011; Manica et al., 2011). The primary role of the PAM may be to maintain a mismatched region between the 5′ tag of the crRNA and the sequence immediately adjacent to the spacer of the target. For DNA targeting systems, this ensures that the chromosomal CRISPR locus is not targeted for cleavage (Marraffini and Sontheimer, 2010b; Mojica et al., 2009). Given that the CMR complex only targets viral RNA, there is no requirement for PAM detection to operate in this case, as the host genome will not be a target. Although no PAM is required, CMR-mediated cleavage requires an unpaired RNA region at the 3′ end of the target RNA, downstream of the protospacer. This discrimination may be at the structural rather than sequence level and is consistent with a role in vivo in targeting viral mRNA sequences, which will typically be considerably longer than the guide RNA species.

The observed activity of SsoCMR differs markedly from that reported previously for the enzyme from *P. furiosus* (Hale et al., 2009). PfuCMR operates by a molecular ruler mechanism without sequence-dependent cleavage, generates 3′-cyclic phosphate products, and does not require an extension at the 3′ end of the target RNA, at least in vitro. The long evolutionary distance between the two species may explain these differences. CMR complexes have been classified into five families (A–E) on the basis of the sequence of the large subunit, Cmr2 (Garrett et al., 2011). On this basis, the SsoCMR complex belongs to family B while PfuCMR is one of the very few representatives of family C. To put these differences in context, the type IIIA, CMR-like complex from *Staphylococcus epidermidis* targets DNA rather than RNA (Marraffini and Sontheimer, 2008). This may therefore reflect the plasticity inherent in the CRISPR system.

The RNA targeting functionality of the CMR complex in prokaryotes has parallels with the eukaryal piRNA pathway that uses guide RNA to recognize and cleave the mRNA of mobile genetic elements (Aravin et al., 2007). As in the CRISPR system, in the piRNA pathway, the small guide RNAs are generated by cleavage of a long mRNA transcript, loaded into an endoribonuclease (Piwi), and used to target and degrade the mobile mRNA by means of dsRNA cleavage, yielding products with 3′-OH termini (reviewed in Nowotny and Yang, 2009). There are, though, important differences. Piwi recognizes the 5′ phosphate of the guide RNA specifically (Ma et al., 2005; Parker et al., 2005), while the guide RNA generated by Cas6 and utilized by CMR lacks a 5′ phosphate and has an essential 8 nt 5′ tag. More fundamentally, there is no obvious homology between the Piwi and CMR proteins. Piwi uses an RNaseH domain for dsRNA cleavage, which is absent from the CMR complex. The stimulation of CMR cleavage activity by ATP suggests that an ATP-driven conformational change may be utilized to reposition the dsRNA with respect to the active site. In the absence of any detectable Walker A or B motifs, the likely site for ATP binding is the C-terminal polymerase/cyclase domain of Cmr2. This domain has no known function but is expected to bind nucleotide triphosphates. If such an ATP-dependent RNA

repositioning mechanism were in operation, it would constitute another aspect of the CMR complex.

Concluding Remarks

In summary, we provide a low-resolution structure of the CMR complex for prokaryotic viral RNA degradation. Deep sequencing suggests that posttranscriptional processing may exert considerable influence on the loading of its crRNA component. The reaction mechanism involves manganese-dependent and ATP-stimulated ribonuclease activity that degrades both target and guide RNA in a sequence-dependent manner. Future studies will aim in particular to map the individual subunits within the EM envelope and the course of the bound crRNA in the complex and to define the function of the polymerase/cyclase domain and the role of ATP in the reaction.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification of Cmr7 Paralogs

Details of cloning, purification, and crystal structure solution for Sso1986 were reported previously (Oke et al., 2010). Sso1725 was expressed and purified according to published protocols (Oke et al., 2010). Briefly, full-length *sso*1725 was cloned into the pDEST14 vector with an N-terminal 6xHis tag and overexpressed in C43 (DE3) *E. coli* at 37°C in LB medium. Expression was induced using 0.4 mM IPTG, and the cultures were harvested after overnight incubation at 25°C. The cell pellets were resuspended and lysed (Soniprep 150, MSE). The lysate was clarified by centrifugation, and then protein was purified by immobilized nickel affinity chromatography and size-exclusion chromatography (Superdex 75 column, GE Healthcare). Sso1725 was concentrated to 10 mg.ml^{-1} for crystallization.

Antibody Generation

Sheep polyclonal antibodies were raised against the recombinant Cmr7 (Sso1986) protein and supplied by the Scottish National Blood Transfusion Service, Pentlands Science Park, Midlothian.

RNA Oligonucleotides Used for CMR Activity Assays

RNA oligonucleotides were chemically synthesized (Integrated DNA Technologies), end labeled with ^{32}P-ATP, and purified by denaturing gel electrophoresis. The sequences used are listed in the Supplemental Experimental Procedures.

Purification of the Native CMR Complex from *S. solfataricus*

S. solfataricus strain P2 biomass was grown as described previously (Götz et al., 2007). The CMR complex was purified over four column chromatography steps, and purification was followed using an antibody raised against subunit Cmr7, as described in the Supplemental Experimental Procedures. This yielded the homogeneous complex shown in Figure 1.

Expression and Purification of Tagged CMR Complex in *S. solfataricus*

This was carried out as described previously by cloning the relevant gene into entry vector pMZ1 (Zolghadr et al., 2007), followed by subcloning into expression vector pSVA9, expressing the relevant subunit with a C-terminal strep-His tag (Albers et al., 2006), described in detail in the Supplemental Experimental Procedures. The Cmr2/Cmr3/Cmr7 subcomplex analyzed by electron microscopy was obtained during purification of the CMR complex with a tagged Cmr3 subunit. The subcomplex eluted separately from the full complex on gel filtration and contained no bound crRNA (Figure S5).

RNA Isolation and Sequencing

RNA was extracted from the purified native CMR complex by the classical phenol/chloroform method followed by ethanol precipitation and vacuum desiccation. Dried RNA was resuspended in 5 μl water and directly labeled in a 10 μl reaction containing polynucleotide kinase and 2 μCi γ^{32}P-ATP. Labeled RNAs were analyzed by electrophoresis on a 15% acrylamide, 7 M urea, TBE denaturing gel and visualized by phosphorimaging. For crRNA deep sequencing, small RNA sequences were generated by the GenePool at the University of Edinburgh using the Illumina small RNA prep kit v1 and subjected to high-throughput sequencing using a Genome Analyzer IIx. This resulted in the addition of the adaptor sequence TCGTATGCCGTCTTCTGCTTG at the 3′ end of each sequence. The adaptor sequence was trimmed away from the reads with a bespoke Perl script. Reads were mapped against the *S. solfataricus* P2 genome with BWA (Li and Durbin, 2009) using default parameters and converted into BAM using SAMtools (Li et al., 2009). Of the 2,527,217 reads, 1,997,151 were mapped (79%). The number and strand orientation of the reads mapping to each spacer were quantified. The raw data from the sequencing run is available from the corresponding author on request, and the sequences mapping onto each spacer are listed in Table S1. The raw sequence data have been uploaded to the Sequence Read Archive with accession number ERP001053 (http://www.ebi.ac.uk/ena/data/view/ERP001053).

RNA Cleavage Assays

Purified SsoCMR complex and unlabeled guide RNA were mixed in buffer (20 mM Mes·HCl [pH 6.0], 100 mM potassium glutamate, 10 mM DTT, 10 mM $MnCl_2$, and the RNase inhibitor SUPERase.In [Ambion]) and preincubated at room temperature for 10 min prior to the addition of 5′-^{32}P-end-labeled synthetic target RNA to the reaction mix. Target and guide RNA and CMR complex concentrations are indicated in the figure legends. The reaction was further incubated at 75°C for 10 min in standard assays or for the time indicated in the figure. Reactions were stopped by chilling on ice and addition of formamide loading buffer. Samples were separated on 20% polyacrylamide, 7 M urea, 1× TBE gels. Electrophoresis was completed at 90 W, 50°C for 90 min, and the gels were visualized by phosphorimaging. 5′ end-labeled RNA size standards (Decade Markers, Ambion) were used to determine the sizes of the observed products. Cas6 activity was assayed as described previously (Lintner et al., 2011).

Crystallography of Sso1725

Sso1725 crystals were grown at 20°C using the sitting drop vapor diffusion method. A reservoir of 0.15 M sodium acetate (pH 5.6), 2.0 M ammonium sulfate was used, and protein was mixed with the precipitant at a ratio of 2:1. Crystals were cryoprotected by successively soaking in solutions of reservoir containing 8%, 16%, 20%, and 25% glycerol before freezing in liquid nitrogen. A native data set at 2.08 Å resolution was collected at Diamond Light Source (Beamline I03). An anomalous SAD data set was collected at the same beamline using a native crystal briefly soaked in reservoir solution containing 42 mM samarium chloride before cryoprotecting as described above. Data sets were processed and refined using the methods described in the Supplemental Experimental Procedures. The coordinates were deposited in the PDB under the accession code 2XVO.

EM Studies

The intact CMR complex bound to crRNA and the Cmr2/Cmr3/Cmr7 subcomplex were studied by negative-staining electron microscopy and single-particle analysis. Data were collected on an FEI F20 FEG microscope, equipped with an 8 k × 8 k CCD camera. Images were collected under low-dose mode at a magnification of 50,000×, at a final sampling of 1.6 Å/pixel at the specimen level. Single-particle images were selected interactively using the Boxer program from the EMAN single-particle analysis package (Ludtke et al., 1999) and extracted into boxes. Image processing was performed using the IMAGIC-5 package (van Heel et al., 1996). The data set was resampled at 6.4 Å/pixel. 10,235 (CMR/RNA) and 5,612 (Cmr2/Cmr3/Cmr7) images were band-pass filtered with a high pass cutoff of 110 Å and a low pass cutoff of 18 Å. The single-particle images were analyzed by Multivariate Statistical Analysis with IMAGIC-5. The data set was subjected to successive rounds of alignment and classification in order to improve the resulting image class averages. Selected CMR/RNA class averages were used to calculate a starting 3D volume by common lines using the Euler program in the IMAGIC-5 package. The CMR/RNA structure was refined until the map converged.

We used the CMR/RNA map to align Cmr2/Cmr3/Cmr7 images and to assign Euler angles by projection matching. Subsequent refinement was carried out until the Cmr2/Cmr3/Cmr7 map converged. Figures were prepared with UCSF Chimera (Goddard et al., 2007).

ACCESSION NUMBERS

The crystal structure of the Cmr7 subunit of CMR (Sso1725) has been submitted to the Protein Data Bank with accession number 2X5Q. The deep sequencing data for the RNA bound to the CMR complex has been uploaded to the Sequence Read Archive with accession number ERP001053.

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, two tables, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.molcel.2011.12.013.

ACKNOWLEDGMENTS

Thanks to Paul Talbot and the University of St. Andrews Mass Spectrometry service for expert technical assistance. This project was funded by the Biotechnology and Biological Sciences Research Council grant number BB/G011400/1. The Electron Microscopy Facility at Edinburgh is supported by the Scottish Alliance for Life Sciences and the Wellcome Trust (WT087658MA). J. Reimann and S.-V.A. were supported by intramural funds of the Max Planck Society.

Received: August 10, 2011
Revised: November 15, 2011
Accepted: December 5, 2011
Published online: January 5, 2012

REFERENCES

Albers, S.V., Jonuscheit, M., Dinkelaker, S., Urich, T., Kletzin, A., Tampé, R., Driessen, A.J., and Schleper, C. (2006). Production of recombinant and tagged proteins in the hyperthermophilic archaeon *Sulfolobus solfataricus*. Appl. Environ. Microbiol. *72*, 102–111.

Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science *318*, 761–764.

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.

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*, 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., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960–964.

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.

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.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J. Bacteriol. *190*, 1390–1400.

Garrett, R.A., Shah, S.A., Vestergaard, G., Deng, L., Gudbergsdottir, S., Kenchappa, C.S., Erdmann, S., and She, Q. (2011). CRISPR-based immune systems of the Sulfolobales: complexity and diversity. Biochem. Soc. Trans. *39*, 51–57.

Goddard, T.D., Huang, C.C., and Ferrin, T.E. (2007). Visualizing density maps with UCSF Chimera. J. Struct. Biol. *157*, 281–287.

Götz, D., Paytubi, S., Munro, S., Lundgren, M., Bernander, R., and White, M.F. (2007). Responses of hyperthermophilic crenarchaea to UV irradiation. Genome Biol. *8*, R220.

Gudbergsdottir, S., Deng, L., Chen, Z., Jensen, J.V., Jensen, L.R., She, Q., and Garrett, R.A. (2011). Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. Mol. Microbiol. *79*, 35–49.

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.

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.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. Science *327*, 167–170.

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J. Bacteriol. *190*, 1401–1412.

Hutvagner, G., and Simard, M.J. (2008). Argonaute proteins: key players in RNA silencing. Nat. Rev. Mol. Cell Biol. *9*, 22–32.

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.

Karginov, F.V., and Hannon, G.J. (2010). The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol. Cell *37*, 7–19.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lillestøl, R.K., Redder, P., Garrett, R.A., and Brügger, K. (2006). A putative viral defence mechanism in archaeal cells. Archaea *2*, 59–72.

Lillestøl, R.K., Shah, S.A., Brügger, K., Redder, P., Phan, H., Christiansen, J., and Garrett, R.A. (2009). CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. Mol. Microbiol. *72*, 259–272.

Lintner, N.G., Kerou, M., Brumfield, S.K., Graham, S., Liu, H., Naismith, J.H., Sdano, M., Peng, N., She, Q., Copié, V., et al. (2011). Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). J. Biol. Chem. *286*, 21643–21656.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. *128*, 82–97.

Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T., and Patel, D.J. (2005). Structural basis for 5′-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature *434*, 666–670.

Makarova, K.S., Aravind, L., Wolf, Y.I., and Koonin, E.V. (2011a). 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. (2011b). Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. *9*, 467–477.

Manica, A., Zebec, Z., Teichmann, D., and Schleper, C. (2011). In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. Mol. Microbiol. *80*, 481–491.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843–1845.

Marraffini, L.A., and Sontheimer, E.J. (2010a). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat. Rev. Genet. *11*, 181–190.

Marraffini, L.A., and Sontheimer, E.J. (2010b). Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature *463*, 568–571.

Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733–740.

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.

Nowotny, M., and Yang, W. (2009). Structural and functional modules in RNA interference. Curr. Opin. Struct. Biol. *19*, 286–293.

Oke, M., Carter, L.G., Johnson, K.A., Liu, H., McMahon, S.A., Yan, X., Kerou, M., Weikart, N.D., Kadi, N., Sheikh, M.A., et al. (2010). The Scottish Structural Proteomics Facility: targets, methods and outputs. J. Struct. Funct. Genomics *11*, 167–180.

Parker, J.S., Roe, S.M., and Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature *434*, 663–666.

Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc. Natl. Acad. Sci. USA *108*, 10098–10103.

Shah, S.A., Hansen, N.R., and Garrett, R.A. (2009). Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. Biochem. Soc. Trans. *37*, 23–28.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. J. Struct. Biol. *116*, 17–24.

Wang, R., Preamplume, G., Terns, M.P., Terns, R.M., and Li, H. (2011). Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. Structure *19*, 257–264.

Wurtzel, O., Sapra, R., Chen, F., Zhu, Y., Simmons, B.A., and Sorek, R. (2010). A single-base resolution map of an archaeal transcriptome. Genome Res. *20*, 133–141.

Zolghadr, B., Weber, S., Szabó, Z., Driessen, A.J., and Albers, S.V. (2007). Identification of a system required for the functional surface localization of sugar binding proteins with class III signal peptides in *Sulfolobus solfataricus*. Mol. Microbiol. *64*, 795–806.