Hot and crispy: CRISPR–Cas systems in the hyperthermophile *Sulfolobus solfataricus*

**Jing Zhang**1,2 and Malcolm F. White1,2

1Biomedical Sciences Research Complex, University of St Andrews, Fife KY16 9ST, U.K.

Abstract

The CRISPR (clustered regularly interspaced short palindromic repeats) and Cas (CRISPR-associated) genes are found in 48% of bacteria and 84% of archaea [1]. The complete sequence of the *Sulfolobus solfataricus* P2 genome revealed six LCTRs (long clusters of tandem repeats) (as they were then known). These had previously been observed in the *Sulfolobus* conjugative plasmid pNOB8, and their function was originally thought to be related to chromosome partitioning [2]. Subsequently, these repeat families, renamed SRSRs (short regularly spaced repeats), were identified in other crenarchaeal genomes [3]. Transcription of *Sulfolobus* SRSR loci to generate small RNAs was first reported in 2005 [4]. Genome analysis revealed the presence of gene families often found adjacent to the repeats, now known as CRISPRs [5]. These Cas proteins were originally thought to constitute a DNA repair system [6], and a subset of these were given the collective name RAMPs (‘repair-associated mysterious proteins’). Subsequently, it was observed that the spacers found between repeats in several organisms including *S. solfataricus* were derived from viral genomes [7–9], leading to the proposal that they constituted an antiviral defence system [10]. The RAMP acronym was revised to stand for ‘repeat-associated mysterious protein’ [10].

Fast-forward 7 years, and the progress made in understanding the mechanisms underlying the CRISPR–Cas system has been astounding. We now understand that CRISPR loci are transcribed to generate a long pre-crRNA (pre-CRISPR RNA) that is processed by a ribonuclease to generate unit length crRNAs (CRISPR RNAs) that are loaded into different types of Cas–protein complexes. These complexes provide immunity against invading genetic elements through homology-directed detection and degradation of foreign nucleic acids in a process known as ‘interference’ (reviewed in [11–14]). Interference complexes have been classified into three major types (I, II and III), with each major type characterized by a unique signature gene [15]. The signature genes for the three types are *cas3, cas9* and *cas10* respectively. Within these three types, CRISPR–Cas systems have been divided further into ten subtypes (I–A–I–F, II–A, II–B, III–A and III–B) [15]. Meanwhile, the acquisition process for new spacers is still not well understood, although the conserved Cas1 and Cas2 proteins play an important role [16,17].

Many bacteria and archaea examined so far have one type of interference complex. Most *Sulfolobus* species, however, harbour complex and diverse CRISPR–Cas systems. In particular, *S. solfataricus* strain P2 has six CRISPR loci (A–F) with repeats belonging to two different families [18], four *cas6* genes and up to seven different interference complexes (two of subtype I-A, three of subtype III-B and one of subtype III-A) (Figure 1). These features make *S. solfataricus* an important model for studying CRISPR–Cas systems.

In the present article, we review our current understanding of crRNA biogenesis and its effector complexes, subtype I-A and subtype III-B, in *S. solfataricus*. We also discuss the differences in terms of mechanisms between the subtype III-B systems in *S. solfataricus* and *Pyrococcus furiosus*.

**Introduction**

The CRISPR (clustered regularly interspaced short palindromic repeats) and Cas (CRISPR-associated) genes are widely spread in bacteria and archaea, representing an intracellular defence system against invading viruses and plasmids. In the system, fragments from foreign DNA are captured and integrated into the host genome at the CRISPR locus. The locus is transcribed and the resulting RNAs are processed by Cas6 into small crRNAs (CRISPR RNAs) that guide a variety of effector complexes to degrade the invading genetic elements. Many bacteria and archaea have one major type of effector complex. However, *Sulfolobus solfataricus* strain P2 has six CRISPR loci with two families of repeats, four *cas6* genes and three different types of effector complex. These features make *S. solfataricus* an important model for studying CRISPR–Cas systems.
Figure 1 | Organization of the genomic CRISPR-Cas locus of *S. solfataricus* strain P2

CRISPR loci are indicated as bold letters (A–F) with the number of spacers indicated. Cas protein effector complexes are boxed and labelled with roman numerals; cas genes are colour-coded according to the key. Four cas6 ribonuclease genes are shown in red and putative transcriptional control proteins of the Csx1/Csm1/Cas3 families are indicated in light green. Selected sso gene numbers are indicated below the genes.

CRISPR loci are indicated as bold letters (A–F) with the number of spacers indicated. Cas protein effector complexes are boxed and labelled with roman numerals; cas genes are colour-coded according to the key. Four cas6 ribonuclease genes are shown in red and putative transcriptional control proteins of the Csx1/Csm1/Cas3 families are indicated in light green. Selected sso gene numbers are indicated below the genes.

**crRNA biogenesis**

CrRNA biogenesis and loading into effector complexes is potentially very complex in *S. solfataricus*. There are four clear Cas6 paralogues that cluster in two distinct families. These may be expressed at different levels and have differing specificities for the two different families of CRISPR repeat that exist in the organism, providing an initial layer of complexity. Furthermore, Cas6 enzymes do not associate stably with the subtype I-A or Type III effector complexes [19,20]. Therefore the determination of factors affecting crRNA cleavage by Cas6 or loading into the effector complexes will be important to understand the molecular mechanism and co-ordination between Cas6 and the complexes.

**Cas6 ribonuclease**

The *S. solfataricus* P2 genome encodes four clear Cas6 proteins. Two of them (Sso1437 and Sso2004; Sso denotes *S. solfataricus*) have been shown to cleave crRNA at 8 nt from the 3’ end [19,21,22]. Analysis of the crystal structure of crRNA-bound Sso2004 revealed that the protein adopts a dimeric arrangement and the RNA adopts a hairpin structure. In contrast, PfuCas6 (Pfu denotes *Pyrococcus furiosus*) contains two tandem RAMP domains that form a groove, in which the repeat RNA wraps around the protein without forming a hairpin structure, analogous to the string around a yo-yo [23]. Although the RNA conformation is dramatically different between the two archaeal crRNA-bound Cas6 proteins, crRNA adopts a similar secondary structure when binding to SsoCas6 and bacterial Cas6 proteins such as *T. thermophilus* and *P. aeruginosa* (reviewed in [24]). These data indicate that SsoCas6 might utilize a crRNA recognition mechanism similar to that in bacteria, but different from that of PfuCas6.

Cas6 is a metal-independent ribonuclease that generates products with 5’-hydroxy and 2’,3’-cyclic phosphate ends, indicating a general acid/base-cleaving mechanism. The catalytic activity requires an essential histidine residue in the active site in almost all Cas6 proteins [25,26]. However, the SsoCas6 and other crenarchaeal Cas6 proteins lack such a histidine residue around the presumed active site, suggesting that these crenarchaeal enzymes might utilize a different mechanism for cleavage. Identification of four conserved basic residues (Lys25, Lys28, Lys51 and Arg231) that are important, but not essential, for catalytic activity began to shed light on this catalytic mechanism [21]. Interestingly, none of these individual residues is absolutely essential for catalysis.

CrRNAs produced by Cas6 programme three types of effector complex in *S. solfataricus* (Figure 2). However, pull-down assays with either Cascade (CRISPR-associated complex for antiviral defence) subunits or CMR (CRISPR-module RAMP) subunits did not lead to co-precipitation of Cas6 proteins [19,20]. The failure suggests that interaction between Cas6 and the CRISPR–Cas complexes, if it exists, is weak and transient. It is therefore more likely that crRNA needs to be released from its processing enzyme and then loaded into its effector complexes after diffusion. This speculation is compatible with a potential for Cas6 to perform multiple rounds of crRNA binding, cleavage and release, unlike the bacterial Cas6 enzymes.

**Subtype I-A (Cascade) complex**

Three operons encoding subtype I-A effector complexes are present in the *S. solfataricus* genome (Figure 1). The gene order is conserved, with each beginning with the *cas5* gene coding for the so-called ‘small’ subunit, followed by the *cas7* and *cas5* genes and then the *cas3*’ and *cas3*’ genes that encode the helicase and nuclease subunits of Cas3. In one gene cluster, there is an extra gene that appears to encode a highly divergent member of the Cas8a2 family (Figure 1). The *Sulfolobus* subtype I-A complex has not been purified from the organism as a defined complex as has been achieved.
in other systems. The Cas7 and Cas5 proteins do form a stable heterodimer that probably constitutes the stable core of all Type I complexes [19], but other subunits interact only weakly. The structure of SsoCas7 was the first for a Cas7 subunit from any CRISPR system, revealing a modified RAMP family fold with a central groove or cleft that was shown to play a role in crRNA binding [19]. The backbone of all Type I (and probably some Type III) complexes is likely to consist of a series of Cas7-like subunits that form a helical structure that serves to bind crRNA [19].

The small subunit, Csa5, is specific to the subtype I-A systems, but recent structural studies have revealed that its fold is related to the C-terminal half of the small subunit of subtype I-E systems (Cse2) [27]. The N-terminal half of Cse2 resembles the fold of the Cmr5 protein, which is the small subunit of the subtype III-B system. Thus the small subunits of Type I and III effector complexes may all be distantly related, as suggested from a bioinformatics analysis [15].

**Figure 2** | Schematic illustration of crRNA biogenesis and effector complex action in *S. solfataricus* P2

---

**Subtype III-B (CMR) complex**

**Protein components and conformation**

*S. solfataricus* P2 genome encodes a subtype III-B CRISPR–Cas system, known as the CMR complex. The complex comprises seven subunits, Cmr1–Cmr7, and a crRNA component, with an overall molecular mass of 430 kDa. Subunits Cmr1–Cmr6 may be present as a single copy in the complex, whereas the Cmr7 subunit is present at a higher stoichiometry. EM structures are available for the full complex and the Cmr2–Cmr3–Cmr7 subcomplex. The full complex displays cavities compatible with an RNA threading machine. The Cmr2–Cmr3–Cmr7 subcomplex contains no crRNA, suggesting a role in RNA binding for Cmr1–Cmr4–Cmr5–Cmr6. Indeed, Cmr1 and Cmr4 are able to bind to RNA *in vitro* (J. Zhang, S. Graham and M.F. White, unpublished work), consistent with the prediction that both proteins have RNA-binding RAMP domains [15]. Determination of the path of RNA in the CMR complex in detail remains a key research direction that should help us to elucidate the catalytic mechanism and subunit organization of the complex.

**Features and profile of crRNA in the complex**

Deep sequencing of RNAs bound in the CMR complex revealed that most were crRNAs, with various lengths that centred on 46 nt. Consistent with the various lengths, some crRNAs have a short 3′ handle derived from the repeat sequence, whereas others have very little. By contrast, crRNA isolated from the subtype I-A complex have the full 3′ handle. This suggests that further maturation of crRNA occurs in the CMR complex, as observed for crRNAs in *P. furiosus* and *Staphylococcus epidermidis* Type III complexes [28,29]. One possible explanation for the contrast is that the *Sulfolobus* CMR complex might recognize and bind to the 5′ tag sequence of crRNA, leaving its free 3′ end susceptible to trimming by an unknown 3′→5′ exonuclease.

**An RNA-dependent ribonuclease**

Although the homology-dependent nuclease activity of the SsoCMR complex has not been demonstrated *in vivo*, the activity has been demonstrated *in vitro* with the purified
whole SsoCMR complex. The SsoCMR complex cleaves target RNA in a sequence-specific manner that is dependent on the presence of a guide crRNA. A perfect match between the crRNAs and their target RNAs, however, is not required for the cleavage reaction, suggesting that the SsoCMR-based immune system might be effective against a variety of virus variants. The cleavage activity of the SsoCMR complex requires manganese and can be stimulated by ATP. The SsoCMR complex cleaves single-stranded target RNAs between UA dinucleotides. The cleavage reaction requires an intact 5' tag sequence in crRNAs. Although cleavage of the crRNA as well as the target RNA molecule could be observed in vitro, one crRNA molecule in the SsoCMR complex can support degradation of multiple molecules of target RNAs [20].

The large subunit Cas10 (also known as Cmr2) contains a putative cyclase-like domain. The domain might bind to Mg\(^{2+}\) and ATP, since the cyclase-like domain in PfuCmr2 is reported to bind to ADP and divalent metals [30]. This binding site may be relevant for the ATP stimulation. Recently, two crystal structures of the complex between Cmr2 and Cmr3 from PfuCMR have revealed a deep channel formed between the subunits. At the foot of this channel lies the most conserved ‘GGDEF’ motif of the cyclase domain. One structure reveals two nucleotides bound in the channel, suggesting an RNA-binding functionality [30]. Together, these data suggest that the cyclase domain may recognize the 5' end of the 8 nt 5' handle of crRNA with the rest of the 5' handle, which does not participate in target RNA binding, bound in the cleft between the Cmr2 and Cmr3 proteins. These predictions require further supporting experimental evidence.

Although both SsoCMR and PfuCMR can carry out crRNA-dependent degradation of target RNAs, there are at least two major differences between the two complexes. First, in PfuCMR, target RNA cleavage site selection uses a ruler mechanism, without discernable sequence dependence [28]. The ruler mechanism may be mediated by the active site in one of the RAMP subunits that is present at multiple copies and forms the backbone of the complex. In contrast, cleavage of RNA by SsoCMR occurs at UA dinucleotides and has no ruler mechanism [20]. Secondly, the cleavage products generated by PfuCMR have 3' phosphate (or 2',3'-cyclic phosphate) and 5' hydroxy ends, whereas SsoCMR produces 5' phosphate and 3' hydroxy ends. These differences may reflect the plasticity and diversity of the CRISPR–Cas systems and suggest that key differences in the nature of the active site in the two systems are likely. Currently, no positive identification of the active site has been reported for either Pyrococcus or Sulfolobus CMR.

Concluding remarks and future directions

Although rapid progress has been made in our understanding of the CRISPR–Cas system in *S. solfataricus*, several key questions remain unanswered:

Have Cas6 paralogues co-evolved with CRISPR repeat sequences?

How are crRNAs loaded into the diverse effector complexes in *S. solfataricus*?

What is the role of the RNA-targeting subtype III-B system in CRISPR–mediated antiviral defence?

How do the three different types of effector complex present in this organism co-operate to defend against invading genetic elements?

Can the acquisition of new spacers be reconstituted in vitro?

These questions will shape our research on the CRISPR system in *S. solfataricus* in the years to come.

### Funding

This work was funded by the Biotechnology and Biological Sciences Research Council [grant numbers BB/J005665/1 and BB/K000314/1 (to M.F.W.)].

### References

8. Pourcel, C., Salvignol, G. and Vergnaud, G. (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151, 653–663

Received 9 July 2013
doi:10.1042/BST20130031