Hot and Crispy: CRISPR-Cas systems in the hyperthermophile *Sulfolobus solfataricus*

Jing Zhang and Malcolm F. White

Biomedical Sciences Research Complex, University of St Andrews, Fife KY16 9ST, UK.
For correspondence: Malcolm White, mfw2@st-andrews.ac.uk; Jing Zhang, jz21@st-andrews.ac.uk

Key words: CRISPR-Cas, Cas6, crRNA, Cascade, CMR, Sulfolobus solfataricus

Abbreviations: CRISPR, cluster of regularly interspaced palindromic repeats; crRNA, CRISPR RNA; RAMP, repeat-associated mysterious protein; CMR, CRISPR module RAMP; Cascade, CRISPR-associated complex for antiviral defence

Summary:
The clustered regularly interspaced palindromic repeat (CRISPR) and CRISPR-associated (Cas) 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 CRISPR RNA (crRNA) that guides 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 repeat, four cas6 genes, and three different types of effector complex. These features make *S. solfataricus* an important model for studying CRISPR-Cas systems. Here, we review our current understanding of crRNA biogenesis and its effector complexes, Type IA and Type IIIB, in *S. solfataricus*. We also discuss the differences in terms of mechanisms between the Type IIIB systems in *S. solfataricus* and *Pyrococcus furiosus*.

Introduction
The cluster of regularly interspersed palindromic repeats (CRISPR) and CRISPR-associated (Cas) genes are found in 48% of bacteria and 84% of archaea [1]. The complete sequence of the *Sulfolobus solfataricus* P2 genome revealed 6 loci of clustered tandem repeats (then known as "long cluster of tandem repeats" (LCTR's)). 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 as Short Regularly Spaced Repeats (SRSR's), were identified in other crenarchaeal genomes [3]. Transcription of *Sulfolobus* SRSR loci to generate small RNA's was first reported in 2005 [4]. Genome analysis revealed the presence of gene families often found adjacent to the repeats, now known as CRISPR's [5]. These CRISPR-associated (Cas) proteins were originally thought to constitute a DNA repair system [6] and a subset of these were given the collective name "Repair Associated Mysterious Proteins" or RAMPs. 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 [10,11]. The RAMP acronym was revised to stand for "Repeat Associated Mysterious Protein" [10].

Fast-forward seven 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-CRISPR RNA (pre-crRNA) that is processed by a ribonuclease to generate unit length crRNAs 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 [12-15]). Interference complexes have been classified into three major types (I, II and III), with each major type characterised by a unique signature gene [16]. The signature genes for the three types are cas3, cas9 and cas10, respectively. Within these three types, CRISPR-Cas systems have been further divided into 10 subtypes (I-A to I-F, II-A, II-B, III-A and III-B) [16]. Meanwhile, the acquisition process for new spacers is still not well understood, although the conserved Cas1 and Cas2 proteins play an important role [17,18].

Many bacteria and archaea examined so far have one type of interference complex. Most *Sulfolobus* species, however, harbour complex and diverse CRISPR-Cas systems [11]. In particular, *Sulfolobus solfataricus* strain P2 has six CRISPR loci (A-F) with repeats belonging to two different families [11], four *cas6* genes and up to seven different interference complexes (2 of Type I-A, 3 of Type III-B and one of Type III-A) (Figure 1). These features make *S. solfataricus* an important model system for the study of CRISPR-Cas immune systems. Hereafter we review the current state of our understanding of *Sulfolobus* crRNA biogenesis and interference.
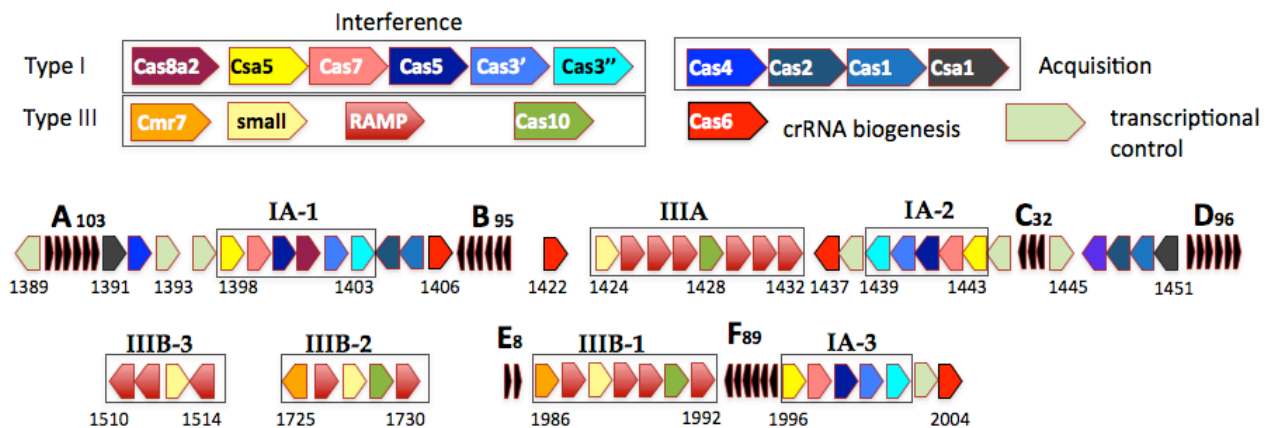

Figure 1. Organisation of the genomic CRISPR-Cas locus of *S. solfataricus* strain P2. CRISPR loci are indicated as bold letters (A-F) with the number of spacers indicated. Cas protein effector complexes are boxed and labelled with roman numerals; *cas* genes are colour coded according to the key. Four *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.

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 Type I-A or Type III effector complexes ([19,20] and unpublished observations). 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 coordination between Cas6 and the complexes.

Cas6 ribonuclease

The *S. solfataricus* P2 genome encodes four clear Cas6 proteins. Two of them (Sso1437 and Sso2004) have been shown to cleave CRISPR repeat RNA at 8 nt from the 3' end [19,21,22]. Analysis of crystal structure of crRNA-bound Sso2004 revealed that the protein adopts a dimeric arrangement and the RNA a hairpin structure. In contrast, PfuCas6 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 yoyo [23]. Although 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 TtCas6e and PaCas6f (reviewed in[24]). These data indicate that SsoCas6 might utilise 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'-hydroxyl 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]. But the SsoCas6 and other crenarchaeal Cas6 proteins lack such a histidine around the presumed active site, suggesting these crenarchaeal enzymes might utilise a different mechanism for cleavage. Identification of four conserved basic residues (K25, K28, K51 and R231) 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). But pull-down assays with either Cascade subunits or CMR 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. Therefore, it is 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.

Type IA – Cascade complex

Three operons encoding type I-A effector complexes are present in the *S. solfataricus* genome (Figure 1). The gene order is conserved, with each beginning with the *csa5* 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* type 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 *S. solfataricus* Cas7 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 type I-A systems, but recent structural studies have revealed that its fold is related to the C-terminal half of the small subunit of type 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 type 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 [16].

Type IIIB (CMR) complex
Protein components and conformation
S. Solfataricus P2 genome encodes a type III-B CRISPR-Cas system, known as the CMR complex. The complex is comprised of seven subunits, Cmr1-7, and a crRNA component, with an overall molecular weight of 430 kDa. Subunits Cmr1-6 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 Cmr2/3/7 subcomplex. The full complex displays cavities compatible with an RNA threading machine. The Cmr2/3/7 subcomplex contains no crRNA, suggesting a role in RNA binding for Cmr1/4/5/6. Indeed, Cmr1 and Cmr4 are able to bind to RNA *in vitro* (authors' unpublished data), consistent with the prediction that both proteins have RNA-binding RAMP domains [16]. Determination of the path of RNA in the CMR complex in

detail remains a key research direction that should help elucidate the catalytic mechanism and subunit organisation 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 variable lengths that centred on 46 nt. Consistent with the variable lengths, some crRNA have a short 3' handle derived from the repeat sequence, while others have very little. By contrast, crRNA isolated from the type IA 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 *S. epidermidis* type III complexes [28,29]. One possible explanation for the contrast is that the *Sulfolobus* CMR complex might recognise and bind to the 5' tag sequence of crRNA, leaving its free 3' end susceptible to trimming by an unknown 3' to 5' exonuclease.
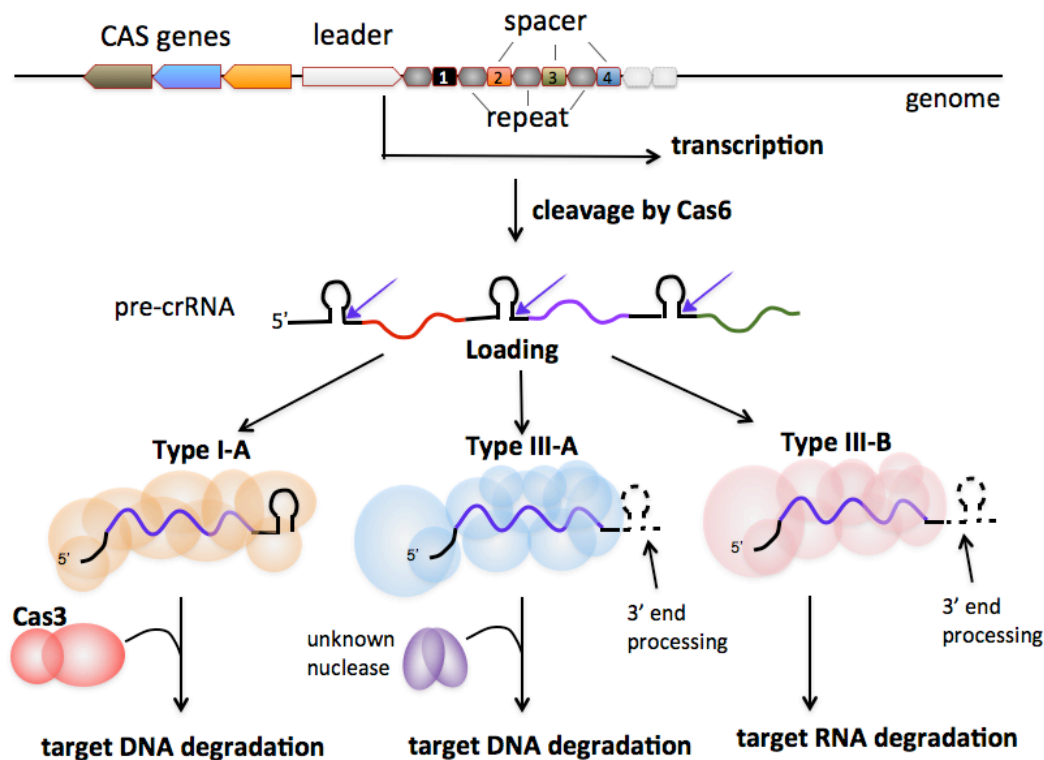

Figure 2. Schematic illustration of crRNA biogenesis and effector complex action in *S. solfataricus* P2.

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 recognise 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 [31]. 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 no ruler mechanism [20]. Second, the cleavage products generated by PfuCMR have 3' phosphate (or 2', 3 cyclic phosphate) and 5' hydroxyl ends, whereas SsoCMR produce 5' phosphate and 3' hydroxyl. 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. Presently, 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 crRNA's loaded into the diverse effector complexes in *S. solfataricus*?
- What is the role of the RNA-targeting type III-B system in CRISPR-mediated antiviral defence?
- How do the three different types of effector complex present in this organism cooperate 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

References:

1   Grissa, I., Vergnaud, G. and Pourcel, C. (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics **8**, 172.

2   She, Q., Phan, H., Garrett, R. A., Albers, S. V., Stedman, K. M. and Zillig, W. (1998) Genetic profile of pNOB8 from Sulfolobus: the first conjugative plasmid from an archaeon. Extremophiles **2**, 417–425.

3   Peng, X., Brügger, K., Shen, B., Chen, L., She, Q. and Garrett, R. A. (2003) Genus-specific protein binding to the large clusters of DNA repeats (short regularly spaced repeats) present in Sulfolobus genomes. Journal of Bacteriology **185**, 2410–2417.

4   Chen, L., Brugger, K., Skovgaard, M., Redder, P., She, Q., Torarinsson, E., Greve, B., Awayez, M., Zibat, A., Klenk, H. P., et al. (2005) The Genome of Sulfolobus acidocaldarius, a Model Organism of the Crenarchaeota. Journal of Bacteriology,

American Society for Microbiology (ASM) **187**, 4992–4999.

5   Jansen, R., Embden, J. D. A. V., Gaastra, W. and Schouls, L. M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. Molecular Microbiology, Blackwell Science Ltd **43**, 1565–1575.

6   Makarova, K. S., Aravind, L., Grishin, N. V., Rogozin, I. B. and Koonin, E. V. (2002) A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. Nucleic Acids Research **30**, 482–496.

7   Mojica, F. J. M., D ez-Villase or, C. S., Garc a-Mart nez, J. S. and Soria, E. (2005) Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. J Mol Evol **60**, 174–182.

8   Pourcel, C. (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.

9   Bolotin, A. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology **151**, 2551–2561.

10 Makarova, K.S., Grishin N.V., Shabalina S.A., Wolf, Y.I. and Koonin, E.V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct, **1**,

7.

11 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. Molecular Microbiology **72**, 259–272.

12 Marraffini, L. A. and Sontheimer, E. J. (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nature Reviews Genetics **11**, 181–190.

13 Barrangou, R. (2013) CRISPR-Cas systems and RNA-guided interference. Wiley Interdiscip Rev RNA **4**, 267–278.

14 Horvath, P. and Barrangou, R. (2010) CRISPR/Cas, the Immune System of Bacteria and Archaea. Science **327**, 167–170.

15 Wiedenheft, B. B., Sternberg, S. H. S. and Doudna, J. A. J. (2012) RNA-guided genetic silencing systems in bacteria and archaea. Nature **482**, 331–338.

16 Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J. M., Wolf, Y. I., Yakunin, A. F., et al. (2011) Evolution and classification of the CRISPR–Cas systems. Nature Publishing Group, Nature Publishing Group **9**, 467–477.

17 Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S. M., Ma, W. and Doudna, J. A. (2009) Structural Basis for DNase Activity of a Conserved Protein Implicated in CRISPR-Mediated Genome Defense. Structure/Folding and Design, Elsevier Ltd **17**, 904–912.

18 Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., et al. (2010) A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Molecular Microbiology **79**, 484–502.

19 Lintner, N. G., Kerou, M., Brumfield, S. K., Graham, S., Liu, H., Naismith, J. H., Sdano, M., Peng, N., She, Q., Copie, V., et al. (2011) Structural and Functional Characterization of an Archaeal Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Complex for Antiviral Defense (CASCADE). Journal of Biological Chemistry **286**, 21643–21656.

20 Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brügger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.-V., et al. (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. Molecular Cell **45**, 303–313.

21 Reeks, J., Sokolowski, R. D., Graham, S., Liu, H., Naismith, J. H. and White, M. F. (2013) Structure of a dimeric crenarchaeal Cas6 enzyme with an atypical active site for CRISPR RNA processing. Biochem J **452**, 223–230.

22 Shao, Y. and Li, H. (2013) Recognition and Cleavage of a Nonstructured CRISPR

RNA by Its Processing Endoribonuclease Cas6. Structure/Folding and Design **21**, 385–393.

23   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/Folding and Design, Elsevier Ltd **19**, 257–264.

24   Reeks, J., Naismith, J. H. and White, M. F. (2013) CRISPR interference: a structural perspective. Biochem J **453**, 155–166.

25   Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. and Doudna, J. A. (2010) Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. Science **329**, 1355–1358.

26   Sashital, D. G., Jinek, M. and Doudna, J. A. (2011) An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. Nature Structural & Molecular Biology, Nature Publishing Group **18**, 680–687.

27   Reeks, J., Graham, S., Anderson, L., Liu, H., White, M. F. and Naismith, J. H. (2013) Structure of the archaeal Cascade subunit Csa5: Relating the small subunits of CRISPR effector complexes. rnabiology **10**, 13–12.

28   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, Elsevier Ltd **139**, 945–956.

29   Hatoum-Aslan, A., Maniv, I. and Marraffini, L. A. (2011) Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. Proceedings of the National Academy of Sciences **108**, 21218–21222.

30   Osawa, T., Inanaga, H. and Numata, T. (2013) Crystal Structure of the Cmr2-Cmr3 Subcomplex in the CRISPR-Cas RNA Silencing Effector Complex. Journal of Molecular Biology **482**, 331–338.

31   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, Elsevier Ltd **139**, 945–956.