The RNA- and DNA-targeting CRISPR–Cas immune systems of *Pyrococcus furiosus*

Rebecca M. Terns*1 and Michael P. Terns*‡1

1Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, U.S.A., †Department of Genetics, University of Georgia, Athens, GA 30602, U.S.A., and ‡Department of Microbiology, University of Georgia, Athens, GA 30602, U.S.A.

Abstract

Using the hyperthermophile *Pyrococcus furiosus*, we have delineated several key steps in CRISPR (clustered regularly interspaced short palindromic repeats)–Cas (CRISPR-associated) invader defence pathways. *P. furiosus* has seven transcriptionally active CRISPR loci that together encode a total of 200 crRNAs (CRISPR RNAs). The 27 Cas proteins in this organism represent three distinct pathways and are primarily encoded in two large gene clusters. The Cas6 protein dices CRISPR locus transcripts to generate individual invader-targeting crRNAs. The mature crRNAs include a signature sequence element (the 5′ tag) derived from the CRISPR locus repeat sequence that is important for function. crRNAs are tailored into distinct species and integrated into three distinct crRNA–Cas protein complexes that are all candidate effector complexes. The complex formed by the Cmr [Cas module RAMP (repeat-associated mysterious proteins)] (subtype III-B) proteins cleaves complementary target RNAs and can be programmed to cleave novel target RNAs in a prokaryotic RNAi-like manner. Evidence suggests that the other two CRISPR-Cas systems in *P. furiosus*, Csa (Cas subtype Apem) (subtype I-A) and Cst (Cas subtype Tneap) (subtype I-B), target invaders at the DNA level. Studies of the CRISPR-Cas systems from *P. furiosus* are yielding fundamental knowledge of mechanisms of crRNA biogenesis and silencing for three of the diverse CRISPR-Cas pathways, and reveal that organisms such as *P. furiosus* possess an arsenal of multiple RNA-guided mechanisms to resist diverse invaders. Our knowledge of the fascinating CRISPR-Cas pathways is leading in turn to our ability to co-opt these systems for exciting new biomedical and biotechnological applications.

Small RNA-guided prokaryotic immunity

To defend themselves from potentially hostile viruses and other invading mobile genetic elements, most archaea and many bacteria rely on small RNA and protein-based CRISPR (clustered regularly interspaced short palindromic repeats)–Cas (CRISPR-associated) adaptive immune systems [1–5]. CRISPR loci consist of tandem arrays of identical repeat sequences interspersed with variable ‘spacer’ sequences acquired from the invaders [6–9]. CRISPR loci are transcribed, and the primary CRISPR RNA transcripts undergo specific processing to generate functionally active mature crRNAs (CRISPR RNAs) [10–13]. Thus CRISPR loci capture virus and plasmid DNA sequences and produce RNAs containing the captured sequences that identify and eliminate the corresponding invaders. The protein components of the immune systems are encoded by specific modules of cas genes, which are typically found near the CRISPR locus [14,15]. Together, the crRNAs and Cas proteins provide RNA-directed immunity against diverse genomic invaders in a wide range of prokaryotic organisms.

Multiple distinct CRISPR–Cas systems have been identified, which differ with regard to cas gene content, CRISPR repeat sequences and mechanistic processes. Ten subtypes of CRISPR–Cas systems were initially identified on the basis of modules of specific cas genes [14,16] and named after an organism that contains only this subtype-specific cas gene cluster (e.g. Cse, *Cas*-subtype *Escherichia coli*) [14]. The CRISPR–Cas systems were recently refined and renamed to recognize categorization into three major types (I–III) (i.e. subtypes I-A to I-F, II-A and II-B, and III-A and III-B, wherein, for example, subtype I-E corresponds to the Cse system) [17]. The distribution of the diverse CRISPR–Cas modules between prokaryotic genomes is highly variable as a result of apparent lateral gene transfer [17,18]. Many bacteria and a few archaea lack CRISPR–Cas systems completely. Moreover, some organisms contain a single CRISPR–Cas system, whereas other organisms, including *Pyrococcus furiosus*, harbour multiple distinct CRISPR–Cas systems.

The CRISPR–Cas systems of *P. furiosus*

*Pyrococcus furiosus* (roughly translated as ‘mad fireball’) is a hyperthermophilic euryarchaeao of the order Thermococcales. *P. furiosus* was initially isolated from thermal marine...
sediments off the coast of Vulcano island, Italy, and exhibits a remarkable optimal growth temperature of \(\sim 100^\circ C\) [19]. Sequenced genomes from six distinct pyrococcal species are available and reveal interesting species-specific diversity in CRISPR–Cas genes and gene organization [20]. We chose *P. furiosus* as a model system to unlock the mysteries of CRISPR–Cas immune systems for several compelling reasons. The inherent thermal stability of the proteins and complexes greatly facilitates detailed mechanistic and structural studies with both recombinant and native complexes. The newly established genetic manipulability of *P. furiosus* [21–23] has further increased the power of *P. furiosus* as a model system and expanded the types of analysis that can be performed. Moreover, as detailed next, *P. furiosus* offers the opportunity to investigate three distinct CRISPR–Cas defence pathways.

The genome of *P. furiosus* strains DSM 3638 and COM1 include seven separate CRISPR loci (distributed throughout the genome) that together encode a total of 200 potential crRNAs (Figures 1A and 1B). All seven CRISPR loci share a common 30 bp repeat sequence interspaced by variable numbers of unique spacers (Figure 1B). The majority of spacer sequences are 37 bp in length, but range in size from 34 to 59 bp. The origins of the spacers are unknown as none map to the few currently known Thermococcales viruses or plasmids [24–27]. The *P. furiosus* repeat RNA is unstructured in solution [28], consistent with the classification of the *P. furiosus* repeat sequence in the ‘unfolded archaeal cluster 6’ family [29]. Adjacent to each CRISPR array is a \(\sim 500\) bp leader region; these are highly conserved among the seven CRISPR loci and contain transcriptional start sites [30] and probably also regulatory elements needed for novel spacer acquisition.

The genes encoding Cas proteins are primarily organized in two large gene clusters in *P. furiosus* (Figures 1A and 1C). The first cas gene cluster encodes the core Cas proteins (grey), the Cmr [Cas module RAMP (repeat-associated mysterious proteins)] proteins (blue) and the Cst (Cas subtype Tneap) proteins (yellow). Csa (Cas subtype Apern) proteins (green) are encoded in the second cas gene cluster. The core cas genes are found in many organisms with different CRISPR–Cas systems and include proteins implicated in new spacer
acquisition (Cas1, Cas2 and Cas4 proteins) as well as crRNA biogenesis (Cas6 protein). Together, the *P. furiosus* genome encodes proteins of three predicted effector immune complexes: Cmr (subtype III-B), Csa (subtype I-A) and Cst (subtype I-B) [14,15]. Additional copies of genes encoding predicted Cmr1, Cas6, Cas4 and three other possible Cas proteins are unlinked and scattered elsewhere in the genome [13] for a total of 27 known or predicted Cas proteins in this organism.

**Expression and processing of crRNAs**

By a combination of deep RNA sequencing and Northern blot analyses, we have determined that all seven *P. furiosus* CRISPR loci are transcribed and give rise to processed mature crRNA species [13,31]. We have mapped the transcription start sites and promoter elements within the CRISPR leader regions [30]. All crRNAs and Cas proteins appear to be expressed in the absence of infection [13,28,30,31], consistent with an immune system operating in ‘surveillance mode’. We also found that crRNAs containing the most recently acquired invader sequences (encoded proximal to the leader) are generally more abundant [13,30,31].

Our studies revealed that *P. furiosus* Cas6 is responsible for primary processing of precursor CRISPR RNA transcripts and plays a critical role in producing individual mature crRNAs [11,28,32] (Figure 2A). We found that Cas6 is a riboendonuclease that makes a single cut within each repeat sequence of the CRISPR transcript, reducing it step-by-step to intermediate RNAs that we call 1× RNAs (Figures 1B and 2A). The 1× RNA intermediates each contain an 8-nt segment of the repeat at the 5′ end (5′-AUUGAAAG-3′) that is retained in the mature crRNAs (the 5′ tag), an invader-targeting or guide sequence, and 22 nt of repeat sequence at the 3′ end (Figure 2A). Cleavage by *P. furiosus* Cas6 produces RNAs with 5′ hydroxy and 2′-3′ cyclic phosphate chemical end groups, consistent with its divalent metal-independent cleavage activity [11,28]. Further 3′ end trimming of 1× crRNA intermediates (by unknown mechanisms) produces the various mature crRNA species (Figures 2A and 2B). Cas6 and Cas6-like enzymes also play a critical role in producing mature crRNAs (and generates 8-nt 5′ tag sequences) in other Type I and III CRISPR–Cas systems [1,10,33–35].

In a series of biochemical and structural studies, we determined how Cas6 specifically recognizes and cleaves precursor crRNAs [11,28,32]. Cas6 recognizes the upstream
The two crRNAs associated with the Crm complex (45 and 39 nt in length) share the common 8-nt 5′ tag (produced by Cas6 cleavage), but have guide sequences (complementary to the target RNA) of 37 and 31 nt respectively, regardless of the length of the guide sequence encoded by the CRISPR spacer [13,30,31]. We reconstituted functional complexes using the recombinant Crm proteins and synthetic crRNAs. Mutational analysis showed that the 5′ tag is critical for the function of the Crm crRNP (CRISPR ribonucleoprotein), probably as a Crm protein recognition site [30]. All six Crm proteins are important for efficient target RNA cleavage (C.R. Hale, R.M. Terns and M.P. Terns, unpublished work).

The structural organization of the *P. furiosus* Crm complex and roles of the individual Crm proteins and crRNA subunits in the target RNA cleavage reaction are currently under investigation. X-ray structures have been obtained for *P. furiosus* Cmr2 [37,38], the Cmr2–Cmr3 heterodimer [39,40] and Cmr5 [41]. Cmr2 belongs to the Cas10 family of Type III CRISPR–Cas system signature proteins [17]. Cmr2 is the largest of the Crm proteins and is composed of several domains including an N-terminal permuted histidine–aspartate (HD) superfamily hydrolase domain, zinc-finger domain, ferredoxin-like fold (cyclase/polymerase domains) and a highly conserved nucleotide/metal-binding domain [37,38]. Cmr2 was a lead candidate for the catalytic component in the Crm complex cleavage, but mutational analysis ruled out a role for both the HD domain and the nucleotide/metal-binding domain of Cmr2 in target RNA cleavage [37]. Cmr5 is the smallest subunit of the complex, is mostly α-helical in nature [41] and associates physically with Cmr4 ([41], and C.R. Hale, R.M. Terns and M.P. Terns, unpublished work). Cmr1, Cmr3, Cmr4 and Cmr6 are predicted RNA-binding proteins as they contain one or two predicted ferredoxin-fold-like domains (cyclase/polymerase domains) and a highly conserved nucleotide/metal-binding domain [37,38]. Cmr2 was a lead candidate for the catalytic component in the Crm complex cleavage, but mutational analysis ruled out a role for both the HD domain and the nucleotide/metal-binding domain of Cmr2 in target RNA cleavage [37].

To date, the Crm effector complex is the only CRISPR–Cas complex known to target RNA (rather than DNA) and is found in ~70% of archaea and ~30% of bacteria with CRISPR–Cas systems, suggesting that it plays an important role in the war against invaders [42]. In recent work, we demonstrated that the Crm complex can be directed to cleave a desired target RNA *in vitro* (in a prokaryotic RNAi-like manner) by programming it with an appropriate crRNA [30], indicating that the site-specific cleavage function of crRNA–Crm complexes could be exploited for gene knockdown.

**Csa and Cst DNA-targeting defence systems**

We have also isolated native *P. furiosus* crRNP complexes containing Cst and Csa proteins and distinct profiles of crRNAs (S. Majumdar, R.M. Terns and M.P. Terns, unpublished work) (Figure 3). We identified the Cas proteins by MS and the crRNAs by deep sequencing and Northern blot analysis. The crRNAs associated with the Cst and Csa proteins all include the 8-nt 5′ repeat sequence tag and

---

**Figure 3** crRNAs are loaded into three distinct crRNP immune effector complexes in *P. furiosus*

Following release by Cas6 and further 3′ end trimming, crRNAs of specific size formed from all seven CRISPR loci become incorporated into the three distinct crRNP complexes containing Crm, Csa or Cst Cas proteins. There is evidence that Crm crRNPs cleave target RNAs, whereas Csa and Cst complexes cleave invader DNA.

The Cmr RNA-targeting defence system

We have isolated and characterized the Cmr effector complex from *P. furiosus* strains DSM 3638 and COM1 (Figure 3). We identified two crRNA species and six Cmr protein subunits associated with native complexes by RNA deep sequencing and MS [13]. The complexes specifically cleave complementary RNAs *in vitro*. The structural organization of the *P. furiosus* Cmr complex and roles of the individual Cmr proteins and crRNA subunits in the target RNA cleavage reaction are currently under investigation. X-ray structures have been obtained for *P. furiosus* Cmr2 [37,38], the Cmr2–Cmr3 heterodimer [39,40] and Cmr5 [41]. Cmr2 belongs to the Cas10 family of Type III CRISPR–Cas system signature proteins [17]. Cmr2 is the largest of the Cmr proteins and is composed of several domains including an N-terminal permuted histidine–aspartate (HD) superfamily hydrolase domain, zinc-finger domain, ferredoxin-like fold (cyclase/polymerase domains) and a highly conserved nucleotide/metal-binding domain [37,38]. Cmr2 was a lead candidate for the catalytic component in the Cmr complex cleavage, but mutational analysis ruled out a role for both the HD domain and the nucleotide/metal-binding domain of Cmr2 in target RNA cleavage [37]. Cmr5 is the smallest subunit of the complex, is mostly α-helical in nature [41] and associates physically with Cmr4 ([41], and C.R. Hale, R.M. Terns and M.P. Terns, unpublished work). Cmr1, Cmr3, Cmr4 and Cmr6 are predicted RNA-binding proteins as they contain one or two predicted ferredoxin-fold-like domains (cyclase/polymerase domains) and a highly conserved nucleotide/metal-binding domain [37,38]. Cmr2 was a lead candidate for the catalytic component in the Cmr complex cleavage, but mutational analysis ruled out a role for both the HD domain and the nucleotide/metal-binding domain of Cmr2 in target RNA cleavage [37]. Cmr5 is the smallest subunit of the complex, is mostly α-helical in nature [41] and associates physically with Cmr4 ([41], and C.R. Hale, R.M. Terns and M.P. Terns, unpublished work). Cmr1, Cmr3, Cmr4 and Cmr6 are predicted RNA-binding proteins as they contain one or two predicted ferredoxin-fold-like domains that are common to many Cas proteins and resemble RRMs (RNA-recognition motifs) found in RNA-binding proteins from all three domains of life [14,15].

To date, the Cmr effector complex is the only CRISPR–Cas complex known to target RNA (rather than DNA) and is found in ~70% of archaea and ~30% of bacteria with CRISPR–Cas systems, suggesting that it plays an important role in the war against invaders [42]. In recent work, we demonstrated that the Cmr complex can be directed to cleave a desired target RNA *in vitro* (in a prokaryotic RNAi-like manner) by programming it with an appropriate crRNA [30], indicating that the site-specific cleavage function of crRNA–Cmr complexes could be exploited for gene knockdown.
possess various 3′ ends that often include a portion of the downstream repeat. Using plasmid challenge assays similar to those used to study the function of the Csa and Cst systems in Thermococcus kodakarenensis in vivo [43], we have obtained evidence that CRISPR–Cas–mediated defences also target DNA in *P. furiosus*, presumably via the action of the Csa and/or Cst effector complexes (J.R. Elmore, R.M. Terns and M.P. Terns, unpublished work).

**Conclusion**

Studies of the CRISPR–Cas systems from *P. furiosus* are yielding fundamental information about mechanisms of crRNA biogenesis and silencing for three of the diverse CRISPR–Cas pathways (Figure 3). These studies reveal that organisms such as *P. furiosus* possess multiple RNA-guided mechanisms (including both RNA- and DNA-targeting systems) that provide layered protection against diverse invaders. Future studies are required to obtain a detailed understanding of how each of the individual RNA- and DNA-targeting systems function and are regulated as well as to determine the extent to which these coexisting systems influence the activity of one another. Detailed knowledge of the CRISPR–Cas pathways is also expected to lead to exciting new biomedical and biotechnological applications.

**Acknowledgements**

We thank the members of the Terns laboratory (present and past) for their contributions to our understanding of *P. furiosus* CRISPR–Cas–based immunity. We express our enthusiastic gratitude to Haruyuki Atomi, Claiborne Glover, Brenton Gravely, Hong Li, Lance Wells and Jan Westpheling for enjoyable and fruitful collaborations.

**Funding**

This work was supported by the National Institutes of Health [grant number R01 GM54682]

**References**


3 Westra, E.R., Swarts, D.C., Staals, R.H., Jore, M.M., Brouns, S.J. and van der Oost, J. (2012) The CRISPRs, they are a-changin’: how prokaryotes acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology **151**, 653–663


9 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

10 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 31 May 2013
doi:10.1042/BST20130056