CRISPR-based immune systems of the Sulfolobales: complexity and diversity

Roger A. Garrett[1], Shiraz A. Shah, Gisle Vestergaard, Ling Deng, Soley Gudbergsdottir, Chandra S. Kenchappa, Susanne Erdmann and Qunxin She

Archaea Centre, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200N Copenhagen K, Denmark

Abstract

CRISPR (cluster of regularly interspaced palindromic repeats)/Cas and CRISPR/Cmr systems of *Sulfolobus*, targeting DNA and RNA respectively of invading viruses or plasmids are complex and diverse. We address their classification and functional diversity, and the wide sequence diversity of RAMP (repeat-associated mysterious protein)-motif containing proteins encoded in Cmr modules. Factors influencing maintenance of partially impaired CRISPR-based systems are discussed. The capacity for whole CRISPR transcripts to be generated despite the uptake of transcription signals within spacer sequences is considered. Targeting of protospacer regions of invading elements by Cas protein–crRNA (CRISPR RNA) complexes exhibit relatively low sequence stringency, but the integrity of protospacer-associated motifs appears to be important. Different mechanisms for circumventing or inactivating the immune systems are presented.

Introduction

The discovery of the widespread occurrence of CRISPR (cluster of regularly interspaced palindromic repeat)-based immune systems in archaea and bacteria has provided important insights into how hosts can inactivate and or regulate invading foreign DNA and, probably, RNA genetic elements. In addition, these systems are likely to influence how co-invading genetic elements can influence one another [1,2]. The two main molecular apparatus involved are structurally complex, partially independent and have diversified functionally. Moreover, their capacity to facilitate the continual uptake of foreign DNA into host chromosomes, and their propensity for transfer between organisms, has important implications for cellular evolution.

The genus *Sulfolobus* provides an important model system for studying these immune systems. Most *Sulfolobus* species carry complex and diverse CRISPR-based systems and appear to be particularly active in the uptake of foreign DNA inserts into their CRISPR loci. Furthermore, a broad collection of *Sulfolobus* genetic elements is available that can be used to challenge the CRISPR-based systems [3]. It includes numerous diverse viruses many of which have been classified into eight new viral families [4,5] as well as a family of plasmids encoding an archaeal-specific conjugative apparatus [6,7].

Many insights into the complexity of the CRISPR-based immune systems, and their mechanistic diversity, have emerged from detailed experimental studies of CRISPR/Cas and CRISPR/Cmr systems of the archaeal genera *Sulfolobus* and *Pyrococcus* respectively, and from investigation of bacterial CRISPR/Cas systems of *Streptococcus thermophilus* [8,9], *Staphylococcus epidermidis* [10,11] and *Escherichia coli* [12]. In the present article, we focus primarily on current knowledge and ideas deriving from, and relating to, the *Sulfolobus* immune systems.

CRISPR/Cas families: complexity, classification and versatility

At an early stage, it was clear that the CRISPR/Cas and Cmr systems were highly complex when approx. 45 different proteins were implicated in their function [13], and the number has continued to rise [14]. Genes of the two systems are clustered into *cas* and *cmr* cassettes which are sometimes linked physically. These cassettes encode a few core proteins, but they also carry different combinations of other genes, some occurring more commonly than others. Thus cassettes vary markedly in their overall gene contents. To illustrate this, core gene structures of the archaeal *cas* cassettes are shown together with a more complex family I *cas* cassette from *Sulfolobus islandicus* HVE10/4 (Figures 1A and 1B). The core *cas* genes classify into *cas* group 1, implicated in CRISPR acquisition of foreign DNA and insertion into CRISPR loci, and *cas* group 2 associated with crRNA (CRISPR RNA) processing and guidance (Figure 1A).

Families of CRISPR/Cas modules have been classified on the basis of gene content and gene order within *cas* cassettes, and on the basis of conserved sequences of *cas* genes, leader regions and repeats within CRISPR/Cas modules. For archaea, about eight families have been proposed, whereas among the Sulfolobales, three are common (I–III) and one less so (IV) [2,15,16,17].

Cmr modules carry two conserved core genes, *cmr2* and *cmr5* (Figure 2A), and a variable number of genes

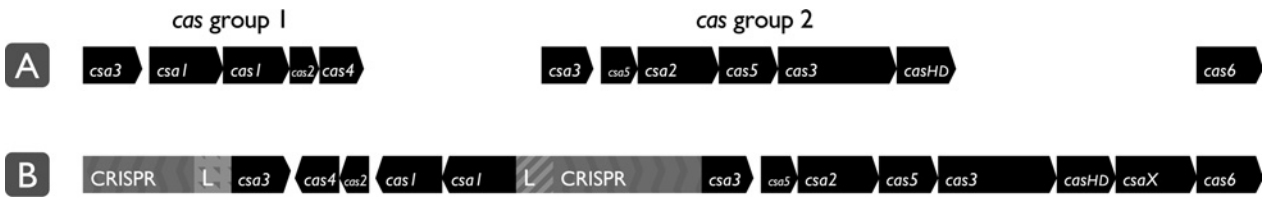
Key words: archaeal virus, cluster of regularly interspaced palindromic repeats/Cas module (CRISPR/Cas module), Cmr module, CRISPR RNA (crRNA), protospacer-associated motif (PAM).

Abbreviations used: CRISPR, cluster of regularly interspaced palindromic repeats; crRNA, CRISPR RNA; IS, insertion sequence; PAM, protospacer-associated motif; RAMP, repeat-associated mysterious protein; SIRV1, *Sulfolobus islandicus* rod-shaped virus 1.

[1]To whom correspondence should be addressed (email garrett@bio.ku.dk).

Figure 1 | Core genes of archaeal *cas* cassettes

(**A**) Core genes are divided into putative functional *cas* groups 1 and 2 (see the text) and the *cas6* gene, which encodes an RNA-processing enzyme [18]. (**B**) Genetic map of a family I CRISPR/Cas module of *S. islandicus* strain HVE10/4 carrying several non-core *cas* genes.



encoding diverse proteins which carry RAMP (repeat-associated mysterious protein) motifs. The Cmr modules can be classified into five main families A, B, C, D and E for archaea on the basis of phylogenetic tree building using sequences of Cmr2 and its homologues Csm1 and Csx11 (Figure 2B), where most *Sulfolobus* Cmr modules fall within families B or D. Further classification is complicated by the presence of multiple diverse copies of genes coding for RAMP-motif-containing proteins. Although these proteins can be classified into families on the basis of these motifs, the remainder of the protein sequences tend to be highly divergent, as illustrated for four proteins encoded in a Cmr family B module of *Sulfolobus solfataricus* P2 (Figure 2C).

Most *Sulfolobus* species carry multiple CRISPR/Cas and/or Cmr modules and, given the high energy cost of maintaining and expressing them, they must confer major advantages on to the cell. Clearly, given the molecular and mechanistic complexities of the systems, they can be inactivated readily by incurring a defect in a component or critical sequence motif. Moreover, the systems are potential targets for incoming genetic elements which may attempt to integrate into essential *cas* or *cmr* genes as has been observed for a viral integration in a *csa3* gene of *S. islandicus* strain M.16.4 (see below) or modify their protein products or otherwise interfere with transcription or maturation of crRNAs. Therefore multiple systems will provide added security against unwanted invasion. The pairing of many family I CRISPR/Cas modules may reflect a compromise between providing added security and generating more compact and efficient systems which can potentially be mobilized and transferred between organisms as single units [2].

A further advantage may arise from the presence of different families of CRISPR/Cas modules which is commonly observed for *Sulfolobus* (e.g. *S. solfataricus* carries family I and II modules, whereas *Sulfolobus acidocaldarius* carries those of family II and III) [16]. Their presence may increase versatility in both the uptake of spacers and targeting of protospacers with different PAMs (protospacer-associated motifs).

The presence of multiple Cmr modules is also likely to confer functional versatility, although they are subject to the constraint that some encoded proteins must be able to recognize part of the repeat sequence of the co-inhabiting CRISPR/Cas module [18,19]. Cmr modules are sometimes linked directly to CRISPR/Cas modules on chromosomes and, given their functional interdependence, there is likely to have been some co-evolution of the coupled systems. Consistent with this view, analysis of the Sulfolobales suggests that Cmr family D modules (Figure 2B) are commonly, but not exclusively, found together with family II CRISPR/Cas modules.

CRISPR loci: structural and functional complexity

CRISPR loci consist of regularly spaced direct repeat sequences with intervening spacers deriving from invading foreign DNA elements. Archaeal repeats fall in the size range 23–37 bp and most spacers are 25–50 bp long [20]. CRISPR loci are preceded by a leader region which varies in size from approx. 150 to 550 bp and shows levels of sequence conservation which are only considered significant within specific families of CRISPR/Cas modules. CRISPR locus sizes can also vary considerably, suggesting that rates of spacer turnover differ markedly for different CRISPR loci within a given archaeon. But there is no support for differences occurring between the CRISPR/Cas families of the Sulfolobales, since large and small clusters exist for the most common families I, II and III.

In organisms carrying several CRISPR/Cas modules, including *S. solfataricus* strains P1 and P2 with six, and *S. acidocaldarius* with five, they may not all be fully functional. The CRISPR/Cas system exhibits two partially independent functions with one group of Cas proteins responsible for uptake of invader DNA into CRISPR loci and the other for generating crRNAs and guiding them to the invading genetic element (Figure 1). Only the latter proteins are essential for the CRISPR/Cas system to function. Thus non-extending CRISPR loci may still be useful to cells as long crRNAs are generated. *S. acidocaldarius* carries two large loci and three smaller ones of 11, five and two spacer-repeat units. All five clusters were transcribed and processed to mature crRNAs [16], but possibly the spacer addition functions are defective for the small clusters. Similarly, for *S. solfataricus* P1 and P2, of the six CRISPR loci, only four appear to be active in elongation. Of the other two, the smallest (locus E) carries six spacer-repeat units with a leader and no *cas* genes [16] and does not appear to be transcribed [21]. It carries spacers matching rudiviruses and a conjugative plasmid and is conserved in three *S. solfataricus* strains (two

Figure 2 | **Classification of archaeal Cmr modules**

(**A**) Gene map of an archaeal Cmr module showing the conserved core proteins Cmr2 and Cmr5, and the grey boxes represent genes encoding different proteins which carry RAMP motifs. (**B**) Phylogenetic tree for archaeal Cmr modules based on the Cmr2 protein sequence showing five main families: A, B, C, D and E. The total number of different proteins in each family carrying RAMP motifs is given in parentheses. Trees were prepared using the MUSCLE and ClustalW programs as described previously [17]. (**C**) Maps of four RAMP motif-containing proteins within a single Cmr family B module of *S. solfataricus* P2. They illustrate the diverse locations of the two conserved amino acid sequence regions (1 and 2), determined using the MEME program [45]. The remaining sequence regions show very low levels of sequence similarity.

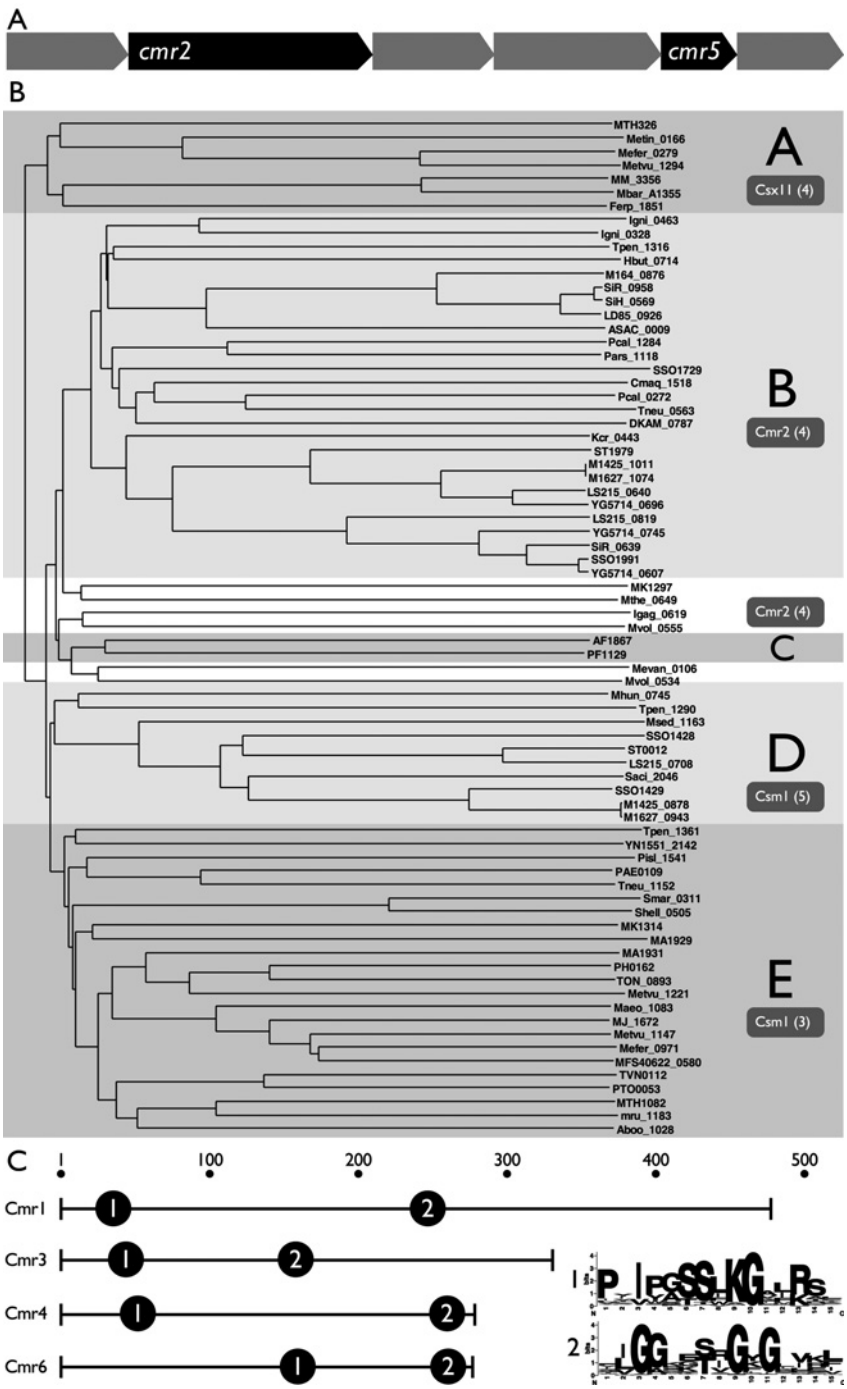
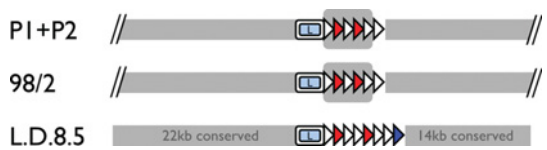


Figure 3 | A map of the CRISPR locus E

Locus E is found in *S. solfataricus* strains P1, P2 and 98/2 and the *S. islandicus* strain L.D.8.5 [22]. Triangles represent spacer-repeat units that are colour-coded for matching sequences: red, rudivirus and blue, conjugative plasmid. The shaded spacer-repeat units carry identical sequences. L represents the leader region. The 36 kb genomic region flanking the locus (grey region) is conserved at >99% sequence identity in all four strains.



from Naples, Italy) with only the final downstream spacer differing between the P1/P2 strains and strain 98/2 (Figure 3). Moreover, it is also found on a highly conserved 36 kb chromosomal fragment (99% sequence identity) in the *S. islandicus* strain L.D.8.5 (from Lassen, CA, U.S.A.) [22], with an almost identical leader region (one mismatch) and identical repeat sequence but different spacers (Figure 3). The maintenance and spreading of locus E, lacking a *cas* cassette, would suggest that the CRISPR module can be activated and generate crRNAs. The inference that Cas proteins encoded in one CRISPR/Cas module can activate other CRISPR loci would also be consistent with the inference that the group 1 *cas* genes (Figure 1A) can exchange between CRISPR/Cas modules [2].

The large inactive locus F with 88 spacer-repeat units, is completely conserved in sequence between *S. solfataricus* strains P1 and P2, but it lacks a leader region, and, although transcription occurs internally within the CRISPR locus, mature crRNAs are not generated [21,23]. Thus the latter, which has been lost from *S. solfataricus* strain 98/2, may be of little use when a viral infection occurred.

Generally for *Sulfolobus* species, loss of mobile DNA elements is difficult, thus IS (insertion sequence) elements tend to degenerate rather than be deleted [24], and this may also apply to CRISPR/Cas and Cmr modules, and explain the maintenance of defective CRISPR systems over long periods, although in a variant strain of *S. solfataricus* P2 (P2A), four physically linked CRISPR/Cas modules (A–D) were apparently lost via a single recombination event between bordering IS elements [25].

Transcription of CRISPR loci and processing

Processed CRISPR transcripts were first observed for the euryarchaeon *Archaeoglobus fulgidus* and crenarchaeon *S. solfataricus*, and these studies revealed the regular pattern of the RNA processing, using probes specific for repeat sequences [26,27]. Subsequently, the smallest *Sulfolobus* RNA product of approx. 40 bp was identified covering primarily a single spacer sequence [20]. *S. acidocaldarius* CRISPR loci are transcribed upstream from the first repeat within the leader region and termination occurs downstream from the final repeat. Even for the locus carrying 78 spacer-repeat units (4930 bp), a substantial proportion of transcripts were approx. 5000 nt long with another large portion in the size range 3000–3500 nt [16].

This raised an important question as to how transcription continues throughout CRISPR loci apparently unimpeded by the presence of spacers carrying archaea-specific promoter or terminator motifs, given that the DNA uptake mechanism is essentially statistically random [15]. A compilation of potential promoter and terminator motifs on the leader (crRNA) strand of the available *Sulfolobus* genomes revealed, for a total of 4505 spacers, 2560 carrying archaeal-type hexameric TATA boxes (at least six consecutive A and Ts with at least two As) and 725 with T-rich pyrimidine motifs (at least six consecutive T and Cs with at least five Ts) [28,29]. Although many of these may at best be weakly effective, nevertheless, given the high gene density in the *Sulfolobus* viral and plasmid genomes and the low frequency of operon structures, the probability of taking up such active motifs is significant. The conclusion that transcripts do not normally start within CRISPR loci is also supported by examination of CRISPR transcripts from *S. solfataricus* P2 transcriptome data [21], which indicate that most of the detectable 5′-ends are attributable to processing sites within repeats [21]. A possible explanation for the unimpeded transcription through the CRISPR loci could be the presence of the CRISPR-binding protein of *Sulfolobus* and other crenarchaea [30]; it could act as a transcription factor inhibiting transcriptional starts and stops within the spacer sequences, and repeats.

Full-length transcripts are also produced from the opposite DNA strand of CRISPR loci of *S. acidocaldarius* which yield discrete 50–60 bp fragments carrying spacer sequences, albeit at lower molar levels than for the crRNAs [16], and antisense RNA transcripts also were detected for CRISPR loci of *S. solfataricus* P2 [21]. Failure to detect similar transcripts in the euryarchaeon *Pyrococcus* and bacterium *E. coli* [12,19] suggests that this may be a specific property of *Sulfolobus* or crenarchaea. Analyses of cDNA libraries of *S. solfataricus* demonstrated previously that antisense RNAs are commonly produced especially against transposase mRNAs [27], and several other antisense RNAs have been detected for this organism [21]. Given that mature crRNAs are produced in the absence of infecting genetic elements in different *Sulfolobus* species [16,20,23], one possible explanation is that these antisense RNAs protect at least a fraction of the crRNAs against degradation before their activation.

Maturation of crRNAs and stringency of targeting mechanisms

Details of RNA-processing mechanism have been elucidated for a euryarchaeal CRISPR/Cmr system and an *E. coli* CRISPR/Cas system where Cas6 homologues cut in the repeat, 8 nt 5′ from the start of the spacer sequence, whereas the 3′-processing sites differ [12,18]. For *S. solfataricus*, many 5′-ends, and putative processing sites, are detectable 6–8 nt from the spacer start [21], suggesting that a similar mechanism

operates. Processing at the 3′-end of the crRNA is less clearly defined, but for the CRISPR/Cmr system of *Pyrococcus*, a 14 nt ruler mechanism enables the processing ribonuclease to generate dual cuts at 5 and 11 nt into the spacer sequence [31]. Presumably, crRNA-binding Cas and Cmr proteins distinguish between the different crRNA products before targeting the foreign DNA or RNA respectively.

Until recently, attention focused on targeting of double-stranded DNA elements, but probably single-stranded DNA will also be targeted by the CRISPR/Cas system. It remains an open question whether the CRISPR/Cmr system targets both mRNA and viral RNA, and incorporation of viral RNA into CRISPR loci would require reverse transcriptase activity. Nevertheless, all evidence suggests that the primary targets of the *Sulfolobus* immune systems are viruses and plasmids and, probably, their mRNAs. There is no support for a general targeting of transposable elements. Spacers matching transposase genes are occasionally found in CRISPR loci [16,20,32], but they can generally be attributed to transposase genes present in viruses or plasmids, in particular orphan *orfB* elements (family IS605/200) for *Sulfolobus* [2,15].

Effective targeting of genetic elements requires that the mature crRNA anneals to the protospacer DNA region. Although, for the bacterium *S. thermophilus*, a perfect sequence match was required to elicit a response from the CRISPR/Cas system [9], studies on different *Sulfolobus* strains have shown that a less stringent recognition system prevails. Challenging *Sulfolobus* cells with viral genes carrying one to three mismatches still produced a strong response from the CRISPR/Cas system [23]. Another important factor is the motif known as PAM. Targeted genetic elements carry this short sequence motif which creates a mismatch with the 5′-end of the crRNA [16,33,34]. For *Sulfolobus*, this was defined as a family-specific dinucleotide, displaced 1 nt from the spacer sequence [15,16]. Potentially, this can be involved in both selection of protospacers for excision by Cas proteins and crRNA targeting. Whereas a study of the bacterium *S. epidermidis* concluded that the PAM was not important for protospacer targeting and that any mismatched base pairing would suffice [11], for *S. islandicus* strain REY15A, altering the PAM led to a loss of crRNA targeting [23].

Anti-immmune systems

Although a few archaeal viruses have been shown to be lytic and to elicit strong immune responses, many *Sulfolobus* viruses and plasmids coexist in a stable relationship, at low copy numbers, over longer periods. Although these genetic elements do not appear to be targeted by the host CRISPR systems, the latter could nevertheless have a regulatory role possibly by targeting mRNAs.

Another special feature of archaeal genetic elements is that they often carry an integrase gene which partitions on chromosomal integration. Consequently, the integrated element can only be excised when the free element is present to generate an intact integrase/excision enzyme [35]. Thus targeting and degradation of the free genetic element by the host CRISPR/Cas system could actually favour entrapment of the integrated element, and such a process could enhance viral and plasmid evolution in archaea. The Redder Model [36] for archaeal viral evolution hypothesized that, since more than one type of fusellovirus can integrate at a given *att* site within a tRNA gene, the encaptured concatenated viruses would tend to recombine thereby generating, and subsequently releasing, hybrid fuselloviruses [36]. A similar process may occur for *Sulfolobus*-specific conjugative plasmids. They are also integrative, and their DNA is regularly incorporated into CRISPR loci as spacers [16,20]. Moreover, this could explain why some of the different Icelandic conjugative plasmids cultivated in Wolfram Zillig's laboratory [37] often carry large regions of almost identical nucleotide sequence [6,7]. Thus, indirectly, the CRISPR/Cas systems could be fuelling production of new viral and plasmid variants which they may subsequently be required to inactivate.

Some insights into how genetic elements undermine or avoid the CRISPR immune systems were gained by passing the rudivirus SIRV1 (*Sulfolobus islandicus* rod-shaped virus 1) through a series of closely related *S. islandicus* strains. This generated many sequence changes in the viral genes, but striking was the frequent occurrence of genes that were altered by 12 bp indels, probably deletions [38]. When similar 12 bp indels were observed among related lipothrixviruses, it was inferred that these might occur at crRNA-targeting protospacers on the viral genomes [39]. In another study of a hyperthermophilic archaeal virus, HAV1 (hyperthermophilic archaeal virus 1), cultured in a bioreactor over a 2-year period, samples taken at different times showed genome sequence changes, not unlike those observed earlier for SIRV1, but also a series of recombination sites were detected along the linear genome at which frequent rearrangements had occurred to generate viral variants with altered sequences [40].

Although accumulating specific sequence changes in genetic elements is an effective way of avoiding, at least temporarily, crRNA targeting, more direct methods must also have evolved. Thus, for the *S. islandicus* strain M.16.4, an M164 provirus 1 has inserted into, and disrupted, the *csa3* gene considered to encode the transcriptional regulator of the group 1 *cas* genes (Figure 1A) associated with new spacer uptake [17]. This has the advantage for the virus that other infecting viruses will still be attacked by crRNAs if matching spacers are already present in the CRISPR locus, but new spacers cannot be generated from M164 provirus itself.

Other possible mechanisms were discerned from a study in which CRISPR systems of *Sulfolobus* were challenged directly by vectors carrying viral genes or protospacers showing various degrees of matching to host CRISPR spacers which mimicked, to a degree, the continual infection of a host cell with a given virus [23]. In many viable transformants, CRISPR locus deletions, including the matching spacer, had occurred, whereas in others, whole CRISPR/Cas cassettes were lost. However, several transformants revealed no changes in either CRISPR/Cas modules or vector constructs,

suggesting that other unknown regulatory mechanisms, can inactivate the immune system [23].

CRISPR/Cas and Cmr module mobility

Sulfolobus CRISPR/Cas and Cmr modules generally occur within variable chromosomal regions where extensive gene shuffling has occurred [2,41], often attributable to high levels of transposable elements. Recombination at bordering IS elements can also lead to loss of CRISPR/Cas or Cmr modules [25]. There is also strong evidence in support of the transfer of whole modules between organisms based on comparative studies of CRISPR/Cas module families and their locations, although the transfer mechanisms remain unclear [2]. For bacteria, evidence was provided for transfer of these modules on large plasmids [42], but many archaeal CRISPR/Cas modules are large, up to 25 kb, and the largest conjugative plasmids are only approx. 40 kb [6]. Chromosomal conjugation may provide a vehicle, possibly facilitated by encaptured *Sulfolobus* conjugative plasmids [43,44] or presently unknown mechanisms may operate, possibly within biofilms. Finally, although phylogenetic analyses support the transfer of CRISPR/Cas and Cmr modules between archaea and bacteria, the basic differences in archaeal and bacterial transcriptional and translational mechanisms and in the unique cell wall, membrane structures and conjugative system of archaea provide formidable barriers to transfer between domains [2].

Funding

Research was supported by grants from the Danish Natural Science Research Council [grant number 272-08-0391], the Danish Research Council for Technology and Production [grant number 274-07-0116] and the Danish National Research Foundation.

References

1 Karginov, F.V. and Hannon, G.J. (2010) The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol. Cell **37**, 7–19
2 Shah, S.A. and Garrett, R.A. (2010) CRISPR/Cas and Cmr modules, mobility and evolution of an adaptive immune system. Res. Microbiol., doi:10.1016/j.resmic.2010.09.001
3 Zillig, W, Arnold, H.P., Holz, I., Prangishvili, D., Schweier, A., Stedman, K., She, Q., Phan, H., Garrett, R. and Kristjansson, J.K. (1998) Genetic elements in the extremely thermophilic archaeon *Sulfolobus*. Extremophiles **2**, 131–140
4 Prangishvili, D., Forterre, P. and Garrett, R.A. (2006) Viruses of the Archaea: a unifying view. Nat. Rev. Microbiol. **11**, 837–848
5 Lawrence, C.M., Menon, S., Eilers, B.J., Bothner, B., Khayat, R., Douglas, T. and Young, M.J. (2009) Structural and functional studies of archaeal viruses. J. Biol. Chem. **284**, 12599–12603
6 Greve, B., Jensen, S., Brügger, K., Zillig, W. and Garrett, R.A. (2004) Genomic comparison of archaeal conjugative plasmids from *Sulfolobus*. Archaea **1**, 231–23
7 Erauso, G., Stedman, K.M., van de Werken, H.J.G., Zillig, W. and van der Oost, J. (2006) Two novel conjugative plasmids from a single strain of *Sulfolobus*. Microbiology **152**, 1951–1968
8 Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science **315**, 1709–1712
9 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
10 Marraffini, L.A. and Sontheimer, E.J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science **322**, 1843–1845
11 Marraffini, L.A. and Sontheimer, E.J. (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature **463**, 568–571
12 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
13 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**, 474–483
14 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
15 Shah, S.A., Hansen, N.R. and Garrett, R.A. (2009) Distributions of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. Biochem. Soc. Trans. **37**, 23–28
16 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
17 Shah, S.A., Vestergaard, G. and Garrett, R.A. (2011) CRISPR/Cas and CRISPR/Cmr immune systems of archaea. In Regulatory RNAs in Prokaryotes (Marchfelder, A. and Hess, W., eds), Springer, Berlin, in the press
18 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
19 Hale, C., Kleppe, K., Terns, R.M. and Terns, M.P. (2008) Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. RNA **14**, 1–8
20 Lillestøl, R.K., Redder, P., Garrett, R.A. and Brügger, K. (2006) A putative viral defence mechanism in archaeal cells. Archaea **2**, 59–72
21 Wurtzel, O., Sapra, R., Chen, F., Zhu, Z.Y., Simmons, B.A. and Sorek, R. (2010) A single-base resolution map of an archaeal transcriptome. Genome Res. **20**, 133–141
22 Reno, M.L., Hel, N.L., Fields, C.J., Burke, P.V. and Whitaker, R.J. (2009) Biogeography of the *Sulfolobus islandicus* pan-genome. Proc. Natl. Acad. Sci. U.S.A. **106**, 8605–8610
23 Gudbergsdottir, S., Deng, L., Chen, Z., Jensen, J.V.K., 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
24 Blount, Z.D. and Grogan, D.W. (2005) New insertion sequences of *Sulfolobus*: functional properties and implications for genome evolution in hyperthermophilic archaea. Mol. Microbiol. **55**, 312–325
25 Redder, P. and Garrett, R.A. (2006) Mutations and rearrangements in the genome of *Sulfolobus solfataricus* P2. J. Bacteriol. **188**, 4198–4206
26 Tang, T.-H., Bachellerie, J.-P., Rozhdestvensky, T., Bortolin, M.-L., Huber, H., Drungowski, M., Elge, T., Brosius, J. and Hüttenhofer, A. (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. Proc. Natl. Acad. Sci. U.S.A. **99**, 7536–7541
27 Tang, T.-H., Polacek, N., Zywicki, M., Huber, H., Brügger, K., Garrett, R.A., Bachellerie, J.P. and Hüttenhofer, A. (2005) Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. Mol. Microbiol. **55**, 469–481

28 Torarinsson, E., Klenk, H.P. and Garrett, R.A. (2005) Divergent transcriptional and translational signals in Archaea. Environ. Microbiol. **7**, 47–54
29 Santangelo, T.J., Cubonová, L., Skinner, K.M. and Reeve, J.N. (2009) Archaeal intrinsic transcription termination *in vivo*. J. Bacteriol. **191**, 7102–7108
30 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. J. Bacteriol. **185**, 2410–2417
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 **139**, 945–956
32 Held, N.L. and Whitaker, R.J. (2009) Viral biogeography revealed by signatures in *Sulfolobus islandicus* genomes. Environ. Microbiol. **11**, 457–466
33 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
34 Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. and Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR system. Microbiology **155**, 733–740
35 She, Q., Peng, X., Zillig, W. and Garrett, R.A. (2001) Gene capture events in archaeal chromosomes. Nature **409**, 478
36 Redder, P., Peng, X., Brügger, K., Shah, S.A., Roesch, F., Greve, B., She, Q., Schleper, C., Forterre, P., Garrett, R.A. and Prangishvili, D. (2009) Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interviral recombination mechanism. Environ. Microbiol. **11**, 2849–2862
37 Prangishvili, D., Albers, S.V., Holz, I., Arnold, H.P., Stedman, K., Klein, T., Singh, H., Hiort, J., Schweier, A., Kristjansson, J.K. and Zillig, W. (1998) Conjugation in archaea: frequent occurrence of conjugative plasmids in *Sulfolobus*. Plasmid **40**, 190–202
38 Peng, X., Kessler, A., Phan, H., Garrett, R.A. and Prangishvili, D. (2004) Multiple variants of the archaeal DNA rudivirus SIRV1 in a single host and a novel mechanism of genomic variation. Mol. Microbiol. **54**, 366–375
39 Vestergaard, G., Shah, S.A., Bize, A., Reitberger, W., Reuter, M., Phan, H., Briegel, A., Rachel, R., Garrett, R.A. and Prangishvili, D. (2008) SRV, a new rudiviral isolate from *Stygiolobus* and the interplay of crenarchaeal rudiviruses with the host viral-defence CRISPR system. J. Bacteriol. **190**, 6837–6845
40 Garrett, R.A., Prangishvili, D., Shah, S.A., Reuter, M., Stetter, K. and Peng, X. (2010) Metagenomic analyses of novel viruses, plasmids, and their variants, from an environmental sample of hyperthermophilic neutrophiles cultured in a bioreactor. Environ. Microbiol. **12**, 2918–2930
41 Brügger, K., Torarinsson, E., Chen, L. and Garrett, R.A. (2004) Shuffling of *Sulfolobus* genomes by autonomous and non-autonomous mobile elements. Biochem. Soc. Trans. **32**, 179–183
42 Godde, J.S. and Bickerton, A. (2006) The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. J. Mol. Evol. **62**, 718–729
43 Aagaard, C., Dalgaard, J. and Garrett, R.A. (1995) Inter-cellular mobility and homing of an archaeal rDNA intron confers selective advantage over intron-cells of *Sulfolobus acidocaldarius*. Proc. Natl. Acad. Sci. U.S.A. **92**, 12285–12289
44 Grogan, D.W. (1996) Exchange of genetic markers at extremely high temperatures in the archaeon *Sulfolobus acidocaldarius*. J. Bacteriol. **178**, 3207–3211
45 Bailey, T.L., Williams, N., Misleh, C. and Li, W.W. (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. **34**, 369–373

Received 30 September 2010
doi:10.1042/BST0390051