The ring of confidence: a haloarchaeal CRISPR/Cas system

Lisa-Katharina Maier*, Britta Stoll*, Jutta Brendel*, Susan Fischer*, Friedhelm Pfeiffer†, Mike Dyall-Smith‡ and Anita Marchfelder*[1]

*Biology II, Ulm University, D-89081 Ulm, Germany, †Department of Membrane Biochemistry, Max-Planck-Institute of Biochemistry, D-52152 Martinsried, Germany, and ‡School of Biomedical Sciences, Charles Sturt University, 2650 NSW, Australia

Abstract

To survive the constant invasions by foreign genetic elements, prokaryotes have evolved various defensive systems. Almost all sequenced archaea, and half of the analysed bacteria use the CRISPR (clustered regularly interspaced short palindromic repeats)/Cas (CRISPR-associated) system, a recently identified prokaryotic immune system that can fend off invading elements in a sequence-specific manner. Few archaeal CRISPR/Cas systems have been analysed so far, and the molecular details of many of the steps involved in adaptation and defence are yet to be understood. In the present paper, we summarize our current knowledge about the CRISPR/Cas system in *Haloferax volcanii*, an extremely halophilic archaeon that was isolated from the Dead Sea. *H. volcanii* encodes a type I-B CRISPR/Cas system, and carries three CRISPR loci and eight Cas proteins. Although in laboratory culture for more than three decades, this defence system was shown to be still active. All three CRISPR loci are transcribed and processed into mature crRNAs (CRISPR RNAs). Cells challenged with engineered plasmids can recognize and eliminate these invading elements if they contain the correct PAM (protospacer adjacent motif) and a sequence that can be recognized by one of the CRISPR spacers.

Prokaryotic defence systems

To defend themselves against foreign genetic elements such as viruses and plasmids, prokaryotes have developed a plethora of strategies [1]. A recently discovered way to fend off invaders is the CRISPR (clustered regularly interspaced short palindromic repeats)/Cas (CRISPR-associated) system (for reviews, see [2–7]). This system provides a ring of confidence that allows prokaryotes to protect themselves against invading genetic elements. Essential for the function of this system are the Cas proteins and the crRNAs (CRISPR RNAs). The latter are transcribed from the CRISPR locus and consist of short sequence repeats in between which spacer sequences are located (Figure 1). The spacer sequences are derived from previous invaders, which have been successfully destroyed. Their DNA has been degraded and a piece of it has been selected to be integrated as a new spacer into the CRISPR locus (Figure 1). Thus the CRISPR locus is a memory of previous attacks which the cell has survived. Since the information about the invader is stored in the genome, the whole system is hereditary. Besides the crRNA, the other key players are the Cas proteins. To date, approximately 65 sets of orthologous Cas proteins have been identified by bioinformatics analyses [8–10]. The Cas proteins have recently been classified into three different major CRISPR/Cas types (I–III), and at least ten subtypes [10]. All CRISPR/Cas systems fend off invader DNA with the exception of type III-B, which degrades invader RNA. Whereas Cas proteins of systems I and III are sufficient to process the CRISPR loci into crRNAs, system II requires a host RNase and a short RNA [tracrRNA (transactivating crDNA)] in addition to the Cas9 protein, in order to generate crRNAs [10]. Systems I and II require a short sequence motif that is encoded in the invader, and lies just next to the sequence which is selected as a new spacer sequence (Figure 1). This motif is termed the PAM (protospacer adjacent motif) [11] and is also required for the recognition of invading elements carrying the same spacer sequence.

The CRISPR/Cas system of *H. volcanii*

The CRISPR/Cas system of *Haloferax* consists of eight Cas proteins and three CRISPR loci (Figure 2A). The *cas* gene cluster encodes the Cas1–Cas8b proteins and is located on minichromosome pHV4, where it is flanked by two of the CRISPR loci (P1 and P2). The third CRISPR locus (locus C) is encoded on the main chromosome. The Cas8b protein is a signature protein for the CRISPR/Cas subtype I-B, classifying the *Haloferax* system as such [12].

All three crRNAs have a repeat sequence of 30 nt in length that is identical except for one residue (Figure 2B). All three RNAs are constitutively expressed and processed, even when cultural conditions such as salt concentration and temperature are varied [12]. This differs from *Escherichia coli*, where crRNA expression is repressed by the histone-like protein H-NS (histone-like nucleoid-structuring protein) [13]. The two crRNAs encoded adjacent to the *cas* gene cluster are as efficiently processed as the crRNA encoded *in trans* on the main chromosome. To identify the invaders

Key words: clustered regularly interspaced short palindromic repeats (CRISPR), CRISPR associated (Cas), crRNA, *Haloferax volcanii*, protospacer adjacent motif (PAM), type I-B.

Abbreviations used: CRISPR, clustered regularly interspaced short palindromic repeats; Cas, CRISPR-associated; crRNA, CRISPR RNA; PAM, protospacer adjacent motif.

[1]**To whom correspondence should be addressed (email anita.marchfelder@uni-ulm.de).**

Figure 1 | **Selection and integration of a new spacer**

Upon entering the cell, the invader DNA is recognized and degraded. A piece of the invader DNA (shown in red) which is adjacent to a PAM sequence (shown in light blue) is selected and integrated into the CRISPR locus as a new spacer. The CRISPR locus consists of the leader sequence (L), which contains the promoter (black arrow), short repeat sequences (shown in black) and spacer sequences (shown in light grey). The new spacer (shown in red) is integrated at the 5′-end of the CRISPR locus.

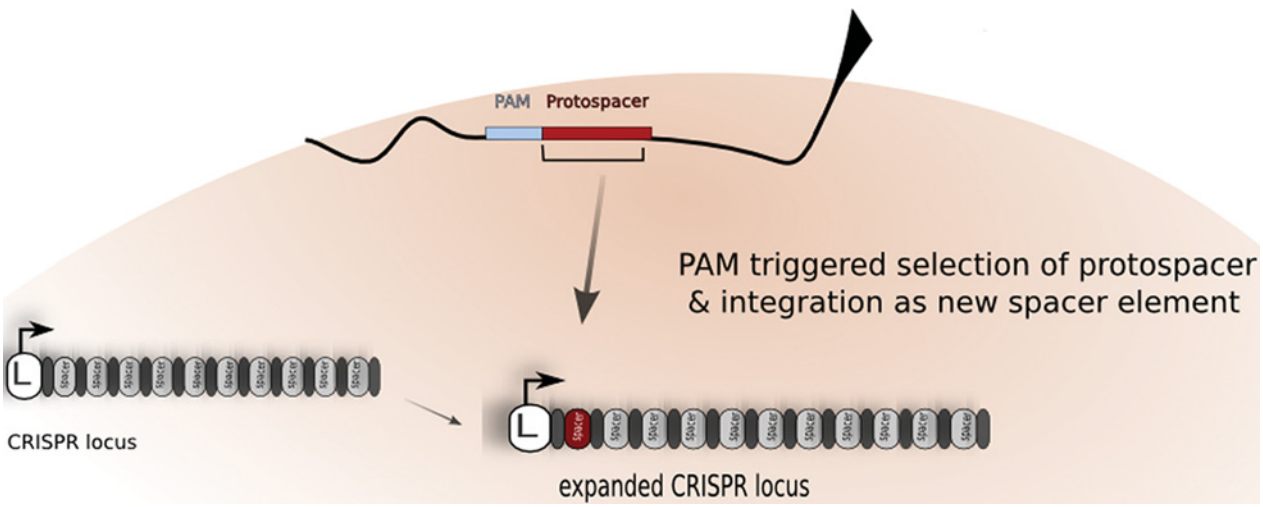
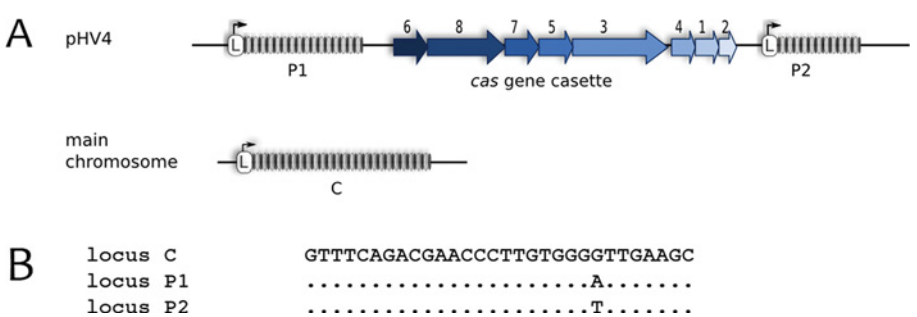


Figure 2 | **The CRISPR/Cas system in *H. volcanii* H119**

(**A**) The *cas* gene cluster is located on the mini-chromosome pHV4 and codes for the proteins Cas1, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7 and Cas8b, with the last-named being the signature protein for subtype I-B. The cluster is flanked by two CRISPR loci (P1 and P2). CRISPR locus P1 contains 17 repeats and 16 spacers (shown in dark grey and light grey respectively), and locus P2 contains 12 repeats and 11 spacers. The third CRISPR locus is encoded on the main chromosome and contains 25 repeats and 24 spacers. The leader region (L) at the 5′-end of the CRISPR locus contains the promoter (black arrow). (**B**) The repeat sequences of the three different crRNAs differ by one residue. The sequence of the CRISPR locus C repeat is shown. Positions that are identical in locus P1 and P2 repeats are displayed with a dot.

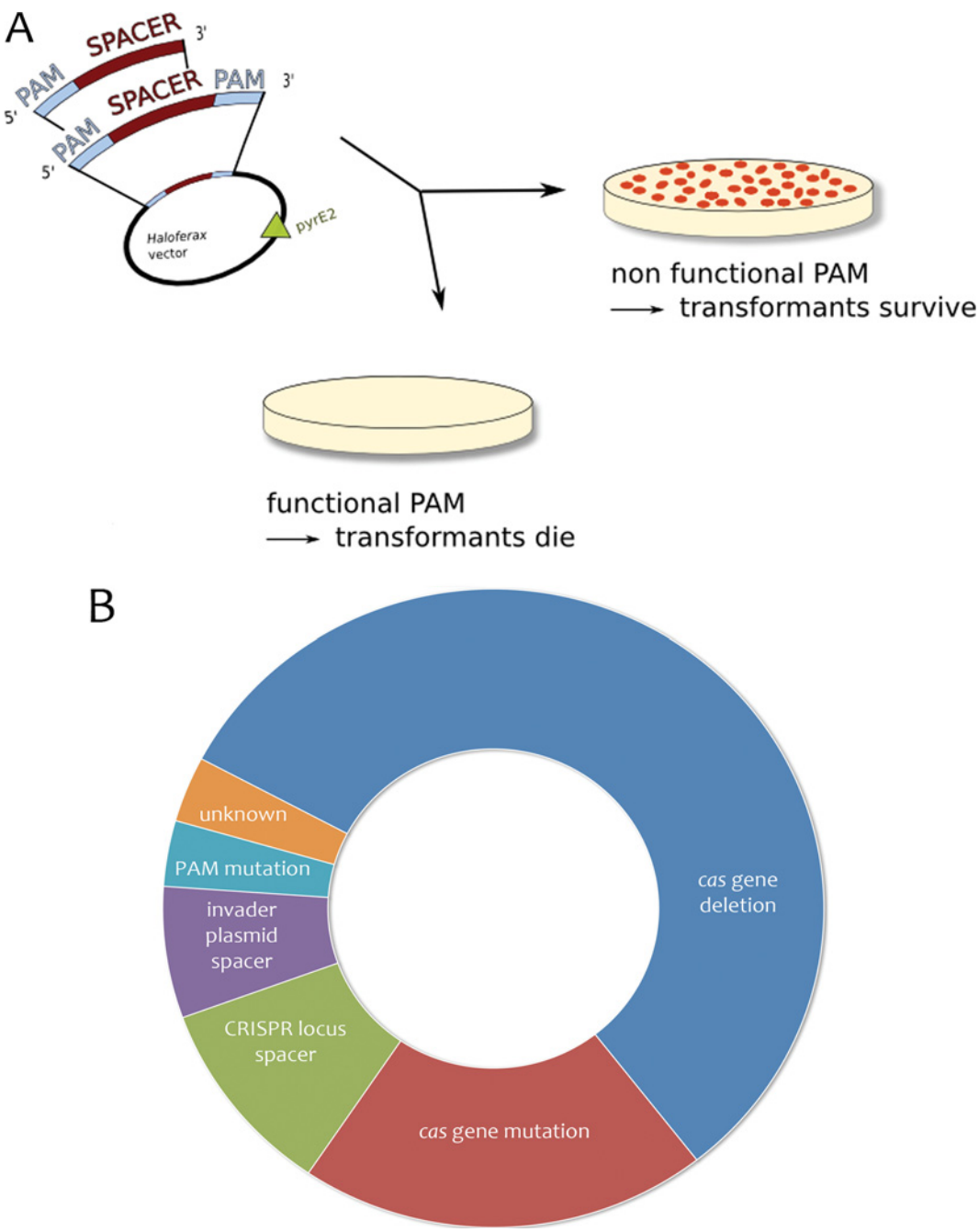


from which the spacers in the *Haloferax* CRISPR loci were derived, all spacer sequences were compared with sequences deposited in public sequence databases. Only two significant matches were found: one against the *Haloferax* genome itself, with an overall similarity of 76%, and another one against an environmental sequence from a sample isolated from an Australian salt lake (Lake Tyrrell), with an overall similarity of 88% [12]. The low number of spacers matches found may be attributed to the fact that the strain was isolated more than 30 years ago, and the virus and plasmid populations may have evolved considerably since then.

To determine the characteristics of the *Haloferax* CRISPR/Cas system, we initiated the analysis of the defence reaction. For this, the crRNA is essential since it recognizes the invader by its sequence. The majority of the *Haloferax* crRNAs were found to consist of spacers that were preceded by 8 nt of the upstream repeat (termed the 5′ handle) and a 3′ trailer consisting of parts of the downstream repeat sequence (L.-K. Maier, B. Stoll, S.J. Lange, J. Brendel, R. Backofen and A. Marchfelder, unpublished work). Similar crRNA structures have been reported in other type I systems [14]. According to frequency analyses, the crRNAs are not present in equal amounts (L.-K. Maier, B. Stoll, S.J. Lange,

Figure 3 | (A) The invader plasmid and (B) mutations that impair the CRISPR/Cas defence

(**A**) To challenge the defence system of *Haloferax,* invader plasmids were constructed which carry the first spacer of the CRISPR locus P1 and trinucleotide sequences adjacent to the spacer as potential PAM sequences. Initially, trinucleotide sequences were cloned on both sides of the spacer until our results showed that the PAM sequence is located upstream. *PyrE2*-deficient *Haloferax* cells were transformed with the 62 different invader plasmids and plated on selection medium. Only cells that retain the plasmid can able to grow on the selection medium. A valid PAM triggers the defence reaction, which then results in degradation of the plasmid and a drastically reduced transformation efficiency (at least 100-fold). (**B**) *H. volcanii* colonies that grew up on selection medium in spite of being transformed with an invader plasmid having a functional PAM were analysed for mutations in the CRISPR/Cas genes. For each clone, the *cas* gene cluster and spacer sequences in the CRISPR locus as well as on the invader plasmid were all sequenced. Most of the 30 clones analysed had deletions in the *cas* genes (57%), and mutations in the *cas* genes were also observed (20%). A few had the spacer sequence in the CRISPR locus P1 deleted (10%), and some had lost the spacer sequence on the invader plasmid (7%). One had the functional PAM sequence mutated (3%) rendering it non-functional, and for one mutant (3%) we did not find any changes, either in the *cas* gene cluster, or in the spacer sequences in the CRISPR locus, or on the invader plasmid.



J. Brendel, R. Backofen and A. Marchfelder, unpublished work).

To investigate which protein(s) are responsible for processing the crRNAs, we generated deletion mutants for each of the eight *cas* genes. Only the deletion of the *cas6* gene resulted in a complete loss of crRNA production (J. Brendel, B. Stoll, L.-K. Maier and A. Marchfelder, unpublished work). Site-directed mutagenesis has been initiated to identify the amino acids important for the Cas6 protein function. Among 21 mutants tested, one resulted in a complete loss of function (J. Brendel, B. Stoll, L.-K. Maier and A. Marchfelder, unpublished work).

The *Haloferax* defence recognizes several PAM sequences

To investigate the prerequisites for the defence reaction, we generated a plasmid-based artificial invader, an experimental approach that has previously been shown to work in *Sulfolobus* [15,16]. For a successful defence reaction, an invader sequence must match a crRNA copy of a spacer sequence, and possess a motif of 2–5 nt that is located adjacent to that sequence in the invader genome (Figure 1). This motif is termed a PAM, and is important for both selection of a protospacer as well as the defence reaction [10,11]. As an artificial invader sequence, we selected the first spacer of CRISPR locus P1. Since the PAM sequences for the *Haloferax* system were not known, we tested all possible trinucleotide combinations for their ability to act as a PAM sequence in this system. It was also not known whether the PAM sequence needed to be located up- or down-stream of the protospacer sequence, so we initially cloned the trinucleotide sequence combinations on both sides of the invader sequence (Figure 3A). A *pyrE2*-deficient *Haloferax* strain was transformed with the invader plasmids, which carried a *pyrE2* gene for selection, allowing growth on uracil-free medium only if cells contained the plasmid. Transformation of 62 different plasmid–invader constructs, each with a distinct potential PAM sequence, showed that six plasmids displayed a greatly reduced (at least 100-fold) transformation rate. These plasmids were specifically recognized and degraded by the defence system based on their match to a CRISPR spacer in combination with the associated functional PAM sequences. The six functional PAM sequences identified are TAA, TAT, TAG, CAC, ACT and TTC. In addition, we could show that the PAM sequence has to be located upstream of the invader sequence. Upon introduction of a functional invader plasmid, not all cells managed to degrade and eliminate the invader, resulting in a small fraction of transformants that were able to grow on the selective medium. These were analysed further and it was found that, in most cases, these cells had suffered deletions or mutations, either in their *cas* genes or in the matching spacer of the CRISPR locus (Figure 3B). Some also carried invader plasmids that had deleted the spacer sequence or had mutations in the PAM. One mutant did not carry mutations or deletions in any of the sequences analysed and is presumed to carry a mutation in some other gene that is required for the defence reaction. Together, these results suggest that the deletion of the *cas* gene cluster is a comparatively frequent event, which raises the question of why the *Haloferax* laboratory strains used in our study (H119 and H26), which do not really need an immune system, have not lost one or more of these genes over the many years of subculture.

PAM interactions

A more detailed analysis of the prerequisites for the defence reaction showed that the crRNA must base pair perfectly at the 5′-end with at least 5 nt (L.-K. Maier, B. Stoll, S.J. Lange, J. Brendel, R. Backofen and A. Marchfelder, unpublished work). A similar kind of seed region has previously been shown to be essential for protospacer recognition by the *E. coli* CRISPR/Cas system, where a contiguous 5 nt seed sequence is also required [17]. In addition, it seems that different spacers have different efficiencies in the defence reaction (L.-K. Maier, B. Stoll, S.J. Lange, J. Brendel, R. Backofen and A. Marchfelder, unpublished work).

Conclusion

We could mimic the invasion of *Haloferax* cells by foreign DNA using a plasmid-based invader carrying a spacer sequence from a *Haloferax* CRISPR locus [12]. Introducing plasmid–invader constructs containing functional PAM sequences resulted in drastically reduced transformation efficiencies, probably due to the complete degradation of the plasmid DNA. This not only confirmed that the defence system of *Haloferax* is active, but also allowed the arrangement and number of functional PAM sequences to be identified in this species. *H. volcanii* was found to have the highest number of such motifs identified so far for a CRISPR repeat group, providing a flexible response that recognizes the original invader as well as sequence variants, and thus increasing the effectiveness of this defence.

In the rare cases of successful invasion, the survivors usually had deletions or mutations of their *cas* genes, so inactivating their entire CRISPR/Cas defence system. This may simply reflect the experimental design, where the invading plasmid allows cell growth but does not carry genes that kill the cell, and neither does it integrate into the chromosome. Even so, one interpretation of this apparently extreme survival mechanism is that it relates to balancing survival of the cell with the risks posed by invading elements. Given that many plasmids and viruses are likely to integrate into the chromosome, a continuing (autoimmune) attack would cause double-strand breaks and probably be lethal. Thus a mechanism to switch off this constitutively expressed defence system would be an advantage.

Acknowledgement

For helpful discussions, we thank all members of the FOR1680 'Unraveling the prokaryotic immune system' Research Unit.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft in the frame of the 'Unravelling the prokaryotic immune system' Research Unit [grant number FOR1680 Ma1538/16-1].

References

1 Labrie, S.J., Samson, J.E. and Moineau, S. (2010) Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. **8**, 317–327
2 Al-Attar, S., Westra, E.R., van der Oost, J. and Brouns, S.J. (2011) Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. Biol. Chem. **2011**, 7
3 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
4 Bhaya, D., Davison, M. and Barrangou, R. (2011) CRISPR–Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu. Rev. Genet. **45**, 273–297
5 Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H. and Moineau, S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature **468**, 67–71
6 Garrett, R.A., Vestergaard, G. and Shah, S.A. (2011) Archaeal CRISPR-based immune systems: exchangeable functional modules. Trends Microbiol. **19**, 549–556
7 Marchfelder, A., Fischer, S., Brendel, J., Stoll, B., Maier, L.K., Jäger, D., Prasse, D., Schmitz, R. and Randau, L. (2012) Small RNAs for defence and regulation in archaea. Extremophiles **16**, 685–696
8 Haft, D.H., Selengut, J., Mongodin, E.F. and Nelson, K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput. Biol. **1**, e60
9 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
10 Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., van der Oost, J. and Koonin, E.V. (2011) Evolution and classification of the CRISPR–Cas systems. Nat. Rev. Microbiol. **9**, 467–477
11 Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. and Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology **155**, 733–740
12 Fischer, S., Maier, L.K., Stoll, B., Brendel, J., Fischer, E., Pfeiffer, F., Dyall-Smith, M. and Marchfelder, A. (2012) An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. J. Biol. Chem. **5**, 5
13 Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N. and Wagner, R. (2010) Identification and characterization of *E. coli* CRISPR–cas promoters and their silencing by H-NS. Mol. Microbiol. **75**, 1495–1512
14 Wiedenheft, B., Sternberg, S.H. and Doudna, J.A. (2012) RNA-guided genetic silencing systems in bacteria and archaea. Nature **482**, 331–338
15 Gudbergsdottir, S., Deng, L., Chen, Z., Jensen, J.V., Jensen, L.R., She, Q. and Garrett, R.A. (2011) Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. Mol. Microbiol. **79**, 35–49
16 Manica, A., Zebec, Z., Teichmann, D. and Schleper, C. (2011) *In vivo* activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. Mol. Microbiol. **80**, 481–491
17 Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J. and Severinov, K. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc. Natl. Acad. Sci. U.S.A. **108**, 10098–10103

Received 15 October 2012
doi:10.1042/BST20120263