The immune system of halophilic archaea

Lisa-Katharina Maier,[1] Susan Fischer,[1] Britta Stoll,[1] Jutta Brendel,[1] Friedhelm Pfeiffer,[2] Mike Dyall-Smith[3] and Anita Marchfelder[1,*]

[1]Biology II; Ulm University; Ulm, Germany; [2]Department of Membrane Biochemistry; Max-Planck-Institute of Biochemistry; Martinsried, Germany; [3]School of Biomedical Sciences; Charles Sturt University; New South Wales, Australia

Prokaryotes have developed several strategies to defend themselves against foreign genetic elements. One of those defense mechanisms is the recently identified CRISPR/Cas system, which is used by approximately half of all bacterial and almost all archaeal organisms. The CRISPR/Cas system differs from the other defense strategies because it is adaptive, hereditary and it recognizes the invader by a sequence specific mechanism. To identify the invading foreign nucleic acid, a crRNA that matches the invader DNA is required, as well as a short sequence motif called protospacer adjacent motif (PAM). We recently identified the PAM sequences for the halophilic archaeon *Haloferax volcanii*, and found that several motifs were active in triggering the defense reaction. In contrast, selection of protospacers from the invader seems to be based on fewer PAM sequences, as evidenced by comparative sequence data. This suggests that the selection of protospacers has stricter requirements than the defense reaction. Comparison of CRISPR-repeat sequences carried by sequenced haloarchaea revealed that in more than half of the species, the repeat sequence is conserved and that they have the same CRISPR/Cas type.

The Prokaryotic Defense System

The CRISPR/Cas system is one of several defense systems that prokaryotes can use to prevent invasion by foreign genetic elements (for a more detailed description see recent reviews[1-6]). The function and significance of this system was only recently discovered, and it differs from other known defense systems because it is heritable, can adapt to new invaders and it is sequence specific. The system uses a set of proteins and short RNA molecules, termed Cas proteins and crRNA, respectively. The crRNAs are processed from a longer pre-crRNA that is encoded in the CRISPR locus, a peculiar series of short, directly repeated sequences between which are unique spacer sequences (**Fig. 1**). The latter sequences originate from previous (and unsuccessful) invading elements, which were degraded. This was accompanied by inserting a short piece of sequence into the CRISPR locus. Thus the CRISPR locus is a memory of previously encountered invaders to which the cell has adapted and is immune.

Immune defense proceeds in three stages: (1) adaptation, (2) expression and (3) interference. In the first stage, the nucleic acid of the invading element enters the cell, and is immediately recognized as a foreign element. A piece of the invader DNA, termed protospacer, is selected and then integrated into the CRISPR locus as a new spacer (**Fig. 2**). Note that the sequence is called protospacer as long as it is still part of the invader. As soon as it is integrated into the CRISPR locus it is called spacer. Selection as a new spacer depends on the presence of a certain neighboring sequence, the protospacer adjacent motif (PAM).[7] This has been shown for CRISPR/Cas type I and type II systems.

This motif is not only important for spacer selection but also for accurately targeting the defense reaction.[7-10]

In the second stage of the defense reaction, the CRISPR locus is expressed, generating a pre-crRNA which is subsequently processed to short crRNAs, each
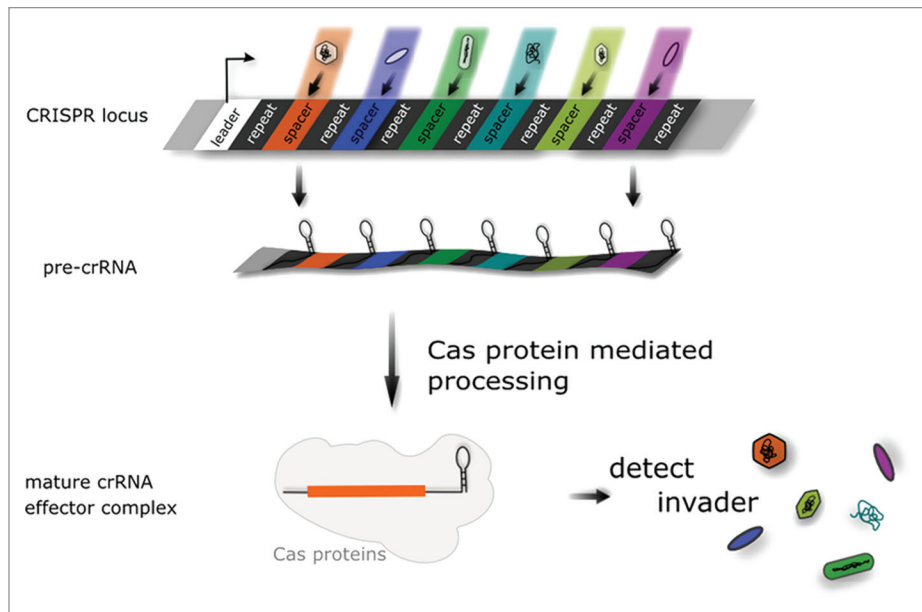
Figure 1. The CRISPR locus. The pre-crRNA is encoded in the CRISPR locus, which consists of repeat (in black) and spacer sequences (colored). In some cases the repeat sequences are able to fold into stem loop structures. The spacer sequences are derived from invader DNA, which previously attacked the cell. CRISPR locus transcription starts from the leader region (black arrow) yielding the pre-crRNA, which is subsequently processed to generate the crRNAs. Each crRNA is specific for one invader.
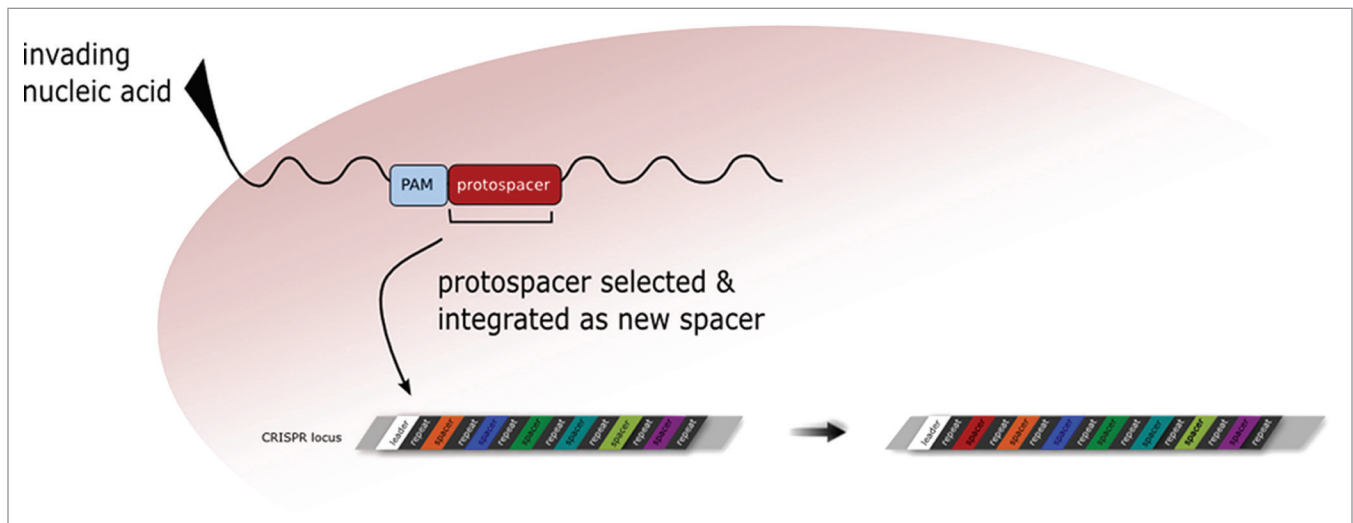

Figure 2. Adaptation to a new invader. Upon entering the cell, the invading foreign nucleic acid is recognized by Cas proteins and a piece of the invader DNA (termed the protospacer, shown in red) is selected to be integrated as a new spacer into the CRISPR locus. A prerequisite to be selected as a new spacer is the presence of the PAM sequence (shown in light blue) adjacent to the protospacer. In Haloferax the PAM sequence has to be located upstream of the protospacer sequence (directly 5′ to it).

of which is specific for a single invader (**Fig. 1**). Together with the Cas proteins, this crRNA recognizes the invader in the third stage of the defense reaction. The spacer sequence of the crRNA base pairs with the invader sequence from which it was derived, rendering the defense sequence specific.

CRISPR/Cas System of *Hfx. volcanii*

The CRISPR/Cas system of Haloferax consists of eight Cas proteins and three CRISPR RNAs, and phylogenetically they belong to the type I-B group of CRISPR/Cas systems.[7,11] We could show

that all three CRISPR RNAs are constitutively expressed and processed,[11] indicating that although the strain has been in the laboratory for more than 30 y and probably did not encounter any invaders during that time, the defense system has remained active. Comparison of the spacer sequences of the three Haloferax

CRISPR loci to sequences deposited in the public sequence databases showed only two matches. One spacer matched to the Haloferax genome within an annotated open-reading frame (HVO_0372) encoding a protein of unknown function. The 5' part of the spacer was identical to the genomic sequence, but the 3' part showed nine mismatches, which is probably sufficient to prevent autoimmune targeting by the CRISPR/Cas system. The second spacer was similar to an environmental sequence recovered from a salt lake in Australia (Lake Tyrrell) and differed at only four positions, distributed along the sequence. The 5' part of the sequence matches perfectly, and in *E. coli* it has been shown that a perfect match in the 5' sequence (termed the seed sequence) is essential for recognition and target degradation.[10,12] It is likely then, that invading elements containing this sequence would be targeted by the CRISPR/Cas system.

The low number of spacer matches to known sequences probably reflects that relatively few haloarchaeal viruses have been isolated and sequenced. Another factor is that the DS2 strain was isolated from the Dead Sea in 1974. Viral populations would have changed in the 40 y since, making it unlikely that the original matching sequences would now be common enough to have been recovered and sequenced.

Several Motifs Direct Degradation

To investigate the Haloferax defense system, we developed a plasmid invader system similar to the one described for Sulfolobus.[9] The invader plasmid contained a piece of invader DNA, and an adjacent motif marking the DNA as invader – the so-called PAM sequence. As invader DNA, we chose a spacer sequence included in the Haloferax CRISPR loci, and this was cloned into a Haloferax shuttle vector (**Fig. 3A**). The PAM sequences used by Haloferax (or by other haloarchaea) were not known before this study, and it was also unknown if they are located upstream or downstream of the protospacer. So we tested all possible di- and trinucleotide combinations (PAM sequences are generally 2–5 nucleotides long) (**Fig. 3A**). In addition, the plasmid contained

a marker gene allowing growth without uracil, so only cells carrying the plasmid are able to grow on selective media. If the defense mechanism is active against the plasmid, then it is destroyed (together with the selection marker) and such cells cannot grow on selective medium, which results in a severe (about 100-fold) reduction of transformation efficiency. Using this approach six different trinucleotide sequences were identified that were active in triggering the defense response, which is currently the highest number of PAMs identified for a single CRISPR repeat group. In addition, we could show that this motif has to be located upstream of the protospacer sequence to activate the defense reaction (**Fig. 3B**).

While the majority of cells challenged with these six types of invader plasmids were unable to grow without uracil, a low level of background colonies were observed. When examples of these were analyzed, the majority was found to have mutations in, or complete deletions of the *cas* gene cluster, thereby inactivating the defense system and allowing the cell to maintain the plasmid.

Conservation of CRISPR/Cas Types in Haloarchaea

To gain more insight into the PAM sequences used for adaptation, we searched for other haloarchaea for which recent metagenomic data were available. We used the spacer sequences encoded in the *Haloquadratum walsbyi* CRISPR loci to look for matches in the databases. Eight matches were found and the PAM sequences obtained for them were in seven cases TTC, which is identical to one of the six PAM sequences we found experimentally for Haloferax. Like Haloferax, *Hqr. walsbyi* contains a CRISPR/Cas type I-B system, with CRISPR repeat sequences that are very similar to those of Haloferax (**Fig. 4**). Further comparison with other haloarchaea showed that those that are available in the CRISPR database (crispr.u-psud.fr/crispr/, July 2012) and which encode Cas proteins all belong to the type I-B CRISPR/Cas group. BLAST searches with the Haloferax repeat sequence show that in 20 of the 32 haloarchaeal genomes currently deposited in
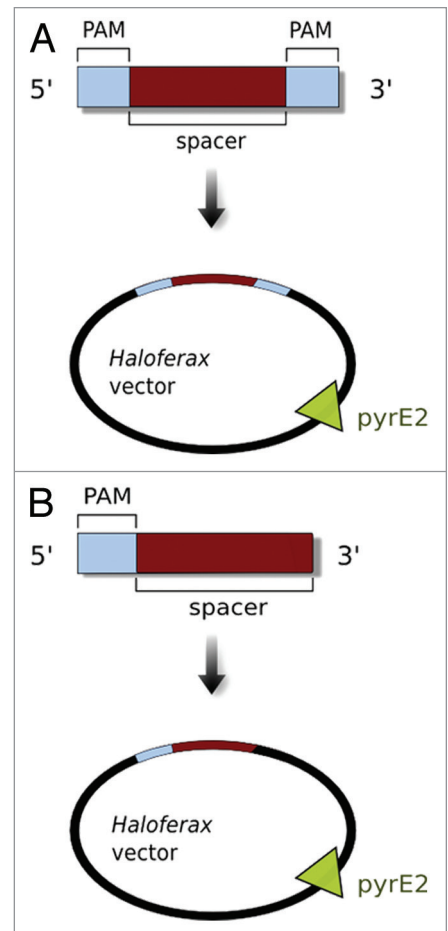

Figure 3. An artificial invader for Haloferax. To challenge the Haloferax defense system, we generated an artificial invader consisting of a spacer sequence (from one of the Haloferax CRISPR loci, shown in red) and an adjacent sequence with all possible two- and three-nucleotide combinations as potential PAM sequences (shown in light blue). These were cloned into a Haloferax plasmid vector that also carried a selection marker *pyrE2* (which makes growth of the Δ*pyrE2-Haloferax* recipient strain independent of supplied uracil). (A) Initial experiments were performed with potential PAM sequences up- and downstream of the spacer sequence. (B) PAM localization experiments showed that the PAM sequence is only required upstream of the spacer sequence and thus we positioned PAM sequences upstream only.

the NBCI database (www.ncbi.nlm.nih.gov/sutils/genom_table.cgi, August 2012) and in the JGI IMG database (http://www.jgi.doe.gov/), the repeat sequence is well conserved, with only one to five mismatches between the repeats from the different haloarchaeal organisms (**Fig. 4**). A similar observation has recently been made by Lynch et al.[13] Since

SPECIES	REPEAT SEQUENCE	Cas type	Proximity
	1 10 20 30		
Hfx. volcanii C	GTTTCAGACGAACCCTTGTGGGGTTGAAGC	I–B	distant
P1	........................A.......	I–B	near
P2	.....................T........	I–B	near
Nmn. pharaonis DSM 2160	....	I–D	distant
Hfx. denitrificans S1	....	I–B	near
Natrinema sp. J7-2	.........................C....	I–B	distant
Har. hispanica ATCC 33960	..................C...........	I–B	near
Nbt. gregoryi SP2	..................C...........	I–B	near
Har. sinaiiensis ATCC 33800	........................A.....	I–B	near
Hfx. sulfurifontis M6	........................A.....	I–D	unknown
Hgm. borinquense DSM 11551	...A..........................	none	–
Halophilic archaeon DL31	..........T...................	I–D	distant
Hmc. katesii DSM 19301	.........G...........A........	I–B	near
Hmc. mukohataei DSM 12286	.........G...........A........	I–B	near
Hrd. tiamatea SARL4B	.........G.....C..............	I–B	near
Hrd. utahensis DSM 12940	.........G.....C..............	I–B	near
Har. californiae BJGN-2	...A.....G.....C..............	I–B	near
Har. marismortui ATCC 43049	...A.....G.....C..............	I–B	near
Hfx. mediterranei ATCC 33500	...A.........A..T.............	I–B	near
Hfx. mucosum PA12	...A.........A..T.............	I–B	near
Halophilic archaeon DL1	..........T.....C...........T	I–B	near
Hqr. walsbyi C23	.........T.........T........T	I–B	near
Nab. magadii ATCC 43099	...............T....A........G	none	–
Hrr. lacusprofundi ATCC 49239	C......C..........C.....T.....	I–B	near

Figure 4. Repeat sequences are conserved in several haloarchaea. The repeat sequence of the *Hfx. volcanii* chromosomally encoded CRISPR locus C differs from the other two CRISPR loci (P1 and P2) by one nucleotide. The repeat sequence of the chromosomally encoded CRISPR locus (locus C) from *Hfx. volcanii* was compared (BLASTN) with the haloarchaeal genomes deposited in the NCBI database (www.ncbi.nlm.nih.gov/sutils/genom_table. cgi, August 2012) and to the JGI IMG database (http://www.jgi.doe.gov/). In 20 of the 32 additionally available genomes at least one CRISPR locus was found where the repeat was conserved, with only one to five mismatches. All of the repeat sequences shown are part of putative CRISPR loci that contained multiple repeats, and were positively identified as CRISPR loci by the CRISPR finder algorithm at (http://crispr.u-psud.fr/). Whether all of these loci represent intact and functional CRISPR/Cas systems is yet to be determined. [a]Distant, a *cas* gene cluster is only found elsewhere in the genome, and is adjacent to another CRISPR locus with a different repeat sequence. [b]Unknown, the genome sequence remains in numerous contigs, and it is not known if a *cas* gene cluster is nearby to the CRISPR locus containing this repeat sequence. After the identification of *cas* gene clusters their proximity to the CRISPR locus was determined by homology searches (e.g., the *cas* gene finder option at CRISPR finder), followed by manual inspection of the annotated genome sequence. Classification of *cas* gene clusters was done according to Makarova et al.[7] *cas* gene clusters are considered near when they are adjacent to the CRISPR locus and distant if found elsewhere in the genome (where they were usually adjacent to another CRISPR locus with a different repeat sequence). The genome of *Hfx. sulfurifontis* remains in numerous contigs, so the proximity of *cas* genes to this locus is currently unknown.

PAM sequences have been reported to be connected to the repeat sequence and to the CRISPR/Cas type[3] it is reasonable to expect that these haloarchaeal sequences require the same PAM sequence. Whether this is true is a question to be addressed in future studies.

Speculation

In our study we analyzed the requirements for the CRISPR/Cas defense reaction and identified six different PAM sequences that were able to trigger this reaction. Such a high number of permissible PAM sequences could be advantageous when defending against related invading elements, as it tolerates individual mutations as well as clonal divergence, making the system more broadly effective. This makes sense because the prokaryotic defense mechanism will remain active against virus mutants that otherwise could avoid immune recognition. In contrast, the few data we collected in silico concerning PAM motifs for *Haloquadratum walsbyi* revealed only two PAM motifs. Since *Hqr. walsbyi* and *Hfx. volcanii* have very similar repeat sequences and belong to the same CRISPR/Cas type they might have similar PAM requirements. Taken together, the evidence suggests that the adaptation step is more restrictive in terms of its PAM requirements compared with the interference step.

In *Streptococcus thermophilus* (CRISPR/Cas type II), a similar recognition of several different PAM sequences on invader plasmids was observed.[2] The PAM requirements of CRISPR/Cas type III systems have not yet been reported, but it has been shown in *Staphylococcus epidermidis*, that the protospacer adjacent sequence of this system must be different from the repeat sequence located upstream of the spacer in the CRISPR locus, so ensuring differentiation between self DNA (CRISPR locus) and the foreign genetic element.[14]

A clearer picture of how this immune reaction operates will be revealed as more data are collected, from all of the known CRISPR/Cas systems.

References

1. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science 2007; 315:1709-12; PMID:17379808; http://dx.doi.org/10.1126/science.1138140.

2. Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 2010; 468:67-71; PMID:21048762; http://dx.doi.org/10.1038/nature09523.

3. Al-Attar S, Westra ER, van der Oost J, Brouns SJ. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. Biol Chem 2011; 392:277-89; PMID:21294681; http://dx.doi.org/10.1515/bc.2011.042.

4. Bhaya D, Davison M, Barrangou R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet 2011; 45:273-97; PMID:22060043; http://dx.doi.org/10.1146/annurev-genet-110410-132430.

5. Garrett RA, Vestergaard G, Shah SA. Archaeal CRISPR-based immune systems: exchangeable functional modules. Trends Microbiol 2011; 19:549-56; PMID:21945420; http://dx.doi.org/10.1016/j.tim.2011.08.002.

6. Marchfelder A, Fischer S, Brendel J, Stoll B, Maier LK, Jäger D, et al. Small RNAs for defence and regulation in archaea. Extremophiles 2012; 16:685-96; PMID:22763819; http://dx.doi.org/10.1007/s00792-012-0469-5.

7. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, et al. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 2011; 9:467-77; PMID:21552286; http://dx.doi.org/10.1038/nrmicro2577.

8. Mojica FJ, Díez-Villaseñor C, García-Martínez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 2009; 155:733-40; PMID:19246744; http://dx.doi.org/10.1099/mic.0.023960-0.

9. Gudbergsdottir S, Deng L, Chen Z, Jensen JV, Jensen LR, She Q, et al. Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. Mol Microbiol 2011; 79:35-49; PMID:21166892; http://dx.doi.org/10.1111/j.1365-2958.2010.07452.x.

10. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, et al. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci U S A 2011; 108:10098-103; PMID:21646539; http://dx.doi.org/10.1073/pnas.1104144108.

11. Fischer S, Maier LK, Stoll B, Brendel J, Fischer E, Pfeiffer F, et al. An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. J Biol Chem 2012; 287:33351-63; PMID:22767603; http://dx.doi.org/10.1074/jbc.M112.377002.

12. Cady KC, Bondy-Denomy J, Heussler GE, Davidson AR, O'Toole GA. The CRISPR/Cas Adaptive Immune System of Pseudomonas aeruginosa Mediates Resistance to Naturally Occurring and Engineered Phages. J Bacteriol 2012; 194:5728-38; PMID:22885297; http://dx.doi.org/10.1128/JB.01184-12.

13. Lynch EA, Langille MG, Darling A, Wilbanks EG, Haltiner C, Shao KS, et al. Sequencing of seven haloarchaeal genomes reveals patterns of genomic flux. PLoS One 2012; 7:e41389; PMID:22848480; http://dx.doi.org/10.1371/journal.pone.0041389.

14. Marraffini LA, Sontheimer EJ. Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature 2010; 463:568-71; PMID:20072129; http://dx.doi.org/10.1038/nature08703.