Chapter 11
CRISPR and Salty: CRISPR-Cas Systems in Haloarchaea

Lisa-Katharina Maier, Omer S. Alkhnbashi, Rolf Backofen, and Anita Marchfelder

Abstract CRISPR-Cas (CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats and Cas: CRISPR associated) systems are unique defence mechanisms since they are able to adapt to new invaders and are heritable. CRISPR-Cas systems facilitate the sequence-specific elimination of invading genetic elements in prokaryotes, they are found in 45% of bacteria and 85% of archaea. Their general features have been studied in detail, but subtype- and species-specific variations await investigation. Haloarchaea is one of few archaeal classes in which CRISPR-Cas systems have been investigated in more than one genus. Here, we summarize the available information on CRISPR-Cas defence in three Haloarchaea: *Haloferax volcanii*, *Haloferax mediterranei* and *Haloarcula hispanica*. Haloarchaea share type I CRISPR-Cas systems, with subtype I-B being dominant. Type I-B systems rely on Cas proteins Cas5, Cas7, and Cas8b for the interference reaction and these proteins have been shown to form a Cascade (CRISPR-associated complex for antiviral defence) -like complex in *Hfx* (*Haloferax*). *volcanii*. Cas6b is the endonuclease for crRNA (CRISPR RNA) maturation in type I-B systems but the protein is dispensable for interference in *Hfx. volcanii*. Haloarchaea share a common repeat sequence and crRNA-processing pattern. A prerequisite for successful invader recognition in *Hfx. volcanii* is base pairing over a ten-nucleotide-long non-contiguous seed sequence. Moreover, *Hfx. volcanii* and *Har* (*Haloarcula*). *hispanica* rely each on certain specific PAM (protospacer adjacent motif) sequences to elicit interference, but they share only one PAM sequence. Primed adaptation in *Har. hispanica* relies on another set of PAM sequences.

Keywords CRISPR-Cas • crRNA • PAM • Cas6 • Cascade • Archaea • *Haloferax volcanii* • *Haloferax mediterranei* • *Haloarcula hispanica* • Type I-B • Adaptation • Interference

L.-K. Maier • A. Marchfelder (✉)
Biology II, University of Ulm, Ulm, Germany
e-mail: anita.marchfelder@uni-ulm.de

O.S. Alkhnbashi • R. Backofen
Bioinformatics Group, Department of Computer Science and Centre for Biological Signalling Studies (BIOSS), University of Freiburg, Freiburg, Germany

B. Clouet-d'Orval (ed.), *RNA Metabolism and Gene Expression in Archaea*,
Nucleic Acids and Molecular Biology 32, DOI 10.1007/978-3-319-65795-0_11

11.1 The Prokaryotic Immune System CRISPR-Cas

Prokaryotic organisms, especially archaea, thrive in nature's most hostile habitats. Apart from the environmental stressors imposed by the abiotic nature of their ecological niche, prokaryotes also face a constant threat by the virome, which exceeds their number by a power of 10 (Suttle 2007). Moreover, prokaryotes face a multitude of invasive entities, including plasmids, transposons and other mobile genetic elements. To balance the integration of beneficial elements and the elimination of detrimental invaders, prokaryotes apply a range of defence strategies (Labrie et al. 2010). The RNA-mediated CRISPR-Cas system has recently attracted increasing attention because it confers adaptive, specific and hereditary immunity against viruses and mobile genetic elements (for recent reviews see (Barrangou 2015; Hille and Charpentier 2016; Mohanraju et al. 2016; Mojica and Rodriguez-Valera 2016)).

Although CRISPR-Cas systems come in different versions (Makarova et al. 2015; Mohanraju et al. 2016; Shmakov et al. 2017), they share one common functional principle. A small RNA guide sequence, called the crRNA, specifically recognizes together with Cas proteins an invading nucleic acid and mediates target degradation. The nature of the effector defines the CRISPR-Cas system as class 1 if a multiprotein complex is present (termed Cascade for type I systems and Csm- or Cmr-complex for type III systems) and class 2 if only a single effector protein is required (e.g. Cas9 for type II, Cas12a[1] for type V) (Makarova et al. 2015; Mohanraju et al. 2016; Shmakov et al. 2017). The *cas* genes include a great variety of nucleic acid binding and processing activities that are crucial for CRISPR-Cas immunity (Jansen et al. 2002a; Makarova et al. 2011). In addition to the mechanistic details of the defence reaction, the presence of these proteins and their characteristic arrangement within the *cas* gene loci give rise to a multilayer classification that currently encompasses 2 classes, 6 types and more than 20 subtypes (Burstein et al. 2017; Makarova et al. 2015; Mohanraju et al. 2016; Shmakov et al. 2015, 2017; Vestergaard et al. 2014).

In contrast the nature of the small RNA guide is relatively uniform. Organisms with an active CRISPR-Cas system encode arrays of recurring repeat sequences that are interspaced by short sequence stretches (spacers) captured from foreign genetic elements in close proximity to the aforementioned *cas* gene cassettes (Bolotin et al. 2005; Jansen et al. 2002a, b; Mojica et al. 2005). The adaptability and expandability of the CRISPR loci through the integration of new spacers of foreign origin upon infection is the basis of the immunogenic power of the CRISPR-Cas system (Barrangou et al. 2007; Brouns et al. 2008; Deveau et al. 2008; Garneau et al. 2010; Pourcel et al. 2005). The cell maintains an ongoing record of previously encountered pathogens or mobile genetic elements that confers specific immunity upon reinfection. Adaptation to an invading genetic element through the integration of new spacers is one of the three stages of CRISPR-Cas immunity (Fig. 11.1)

[1]Cas12a was formerly termed Cpf1.

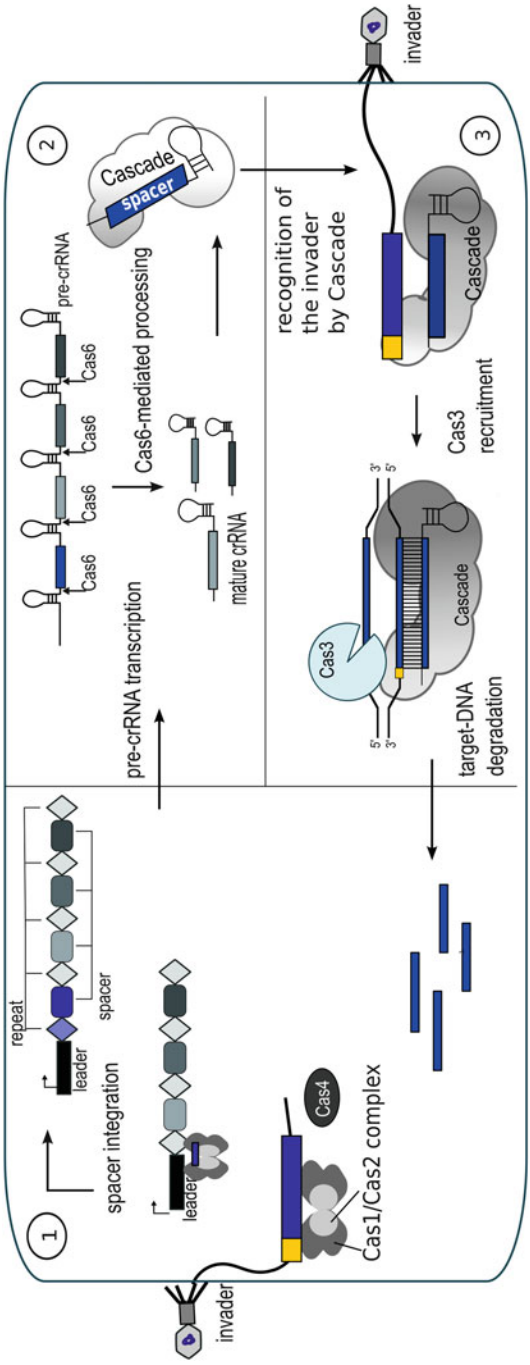

Fig. 11.1 The three stages of type I interference. The first stage (1) ensures the recognition of a foreign invader via a Cas1/Cas2 complex (adaptation step). Part of the invading DNA, the protospacer sequence, is selected by identifying the PAM sequence. Cas1/Cas2-mediated integration into the CRISPR locus at the leader end expands the array by one repeat spacer unit. All type I-B systems contain also a Cas4 protein, that has been shown to be involved in adaptation in *Har. hispanica* (Li et al. 2014b). However, details on how Cas4 is involved in the adaptation step are not known yet. The transcription of the CRISPR locus into a long pre-crRNA initiates the second stage (2) (expression stage). After being processed into mature crRNAs by the Cas6 endonuclease, the crRNAs are integrated into the multiprotein effector complex termed Cascade. The crRNA-loaded Cascade complexes can detect the foreign invader during the final interference stage (3). Upon reinfection with a cognate invader, Cascade scans the foreign nucleic acid, and upon identification of the PAM sequence and base pairing of the crRNA spacer region and the protospacer sequence, the nuclease Cas3 is recruited and the invader DNA is degraded

(detailed reviews can be found in (Amitai and Sorek 2016; Mohanraju et al. 2016; Sternberg et al. 2016)). To use the genetic information stored within the spacer sequences, the CRISPR array is transcribed in the second stage of the defence reaction into a long precursor molecule, the pre-crRNA, which is subsequently processed into the mature crRNAs. This reaction is catalysed in type I systems by Cas proteins and in most type II systems by RNase III in conjunction with tracrRNA and Cas9. Every crRNA comprises a unique spacer flanked by the remainder of the repeat sequence. Depending on the type of system, each crRNA is joined by one or more Cas proteins to form the active effector complex, in type II and type V-B systems the effector complex also contains the tracrRNA (Shmakov et al. 2017). In the third and final step, the interference, the crRNA-loaded complex mediates the recognition of foreign nucleic acid sequences through the base pairing between the crRNA and the invader. This interaction leads to the degradation and subsequent elimination of the targeted nucleic acid: in type I systems the degrading nuclease Cas3 is recruited to the effector complex whereas in type III systems Cas10 and Cmr/Csm subunits of the effector complex mediate target degradation; in class 2 systems degradation is achieved via the activity of the single effector protein. Type I, II, and V systems target DNA, whereas the activity of type III systems is transcription-dependent and results in degradation of RNA and DNA (Mohanraju et al. 2016), the recently discovered type VI systems target RNA (Mohanraju et al. 2016; Shmakov et al. 2017).

Moreover, some CRISPR-Cas types rely on short sequence motifs, termed PAM (type I, II and V) or PFS[2] (type VI), that are located on the invading nucleic acid (Deveau et al. 2008; Jinek et al. 2012; Mojica et al. 2009; Shmakov et al. 2017; Westra et al. 2013; Zetsche et al. 2015). These motifs direct protospacer selection during adaptation as well as interference by ensuring self/non-self discrimination (Amitai and Sorek 2016; Shah et al. 2013).

Despite the rapid pace in scientific activity regarding CRISPR-Cas systems, much has yet to be elucidated with regard to the protein and RNA machinery that execute CRISPR-Cas function as well as the regulatory circuits that orchestrate it.

11.2 CRISPR-Cas Systems in Haloarchaea

CRISPR-Cas systems are present in approximately 45% of bacteria and 85% of archaea (Alkhnbashi et al. 2014; Lange et al. 2013; Makarova et al. 2015). Despite their prevalence in archaea, most studies have focused on CRISPR-Cas systems in bacteria, whereas only a few archaeal model organisms have been analysed. Archaeal CRISPR-Cas systems are almost exclusively restricted to class 1 systems, which rely on a multisubunit effector complex (Makarova et al. 2015; Vestergaard et al. 2014). Just recently a few class 2 systems have been found in archaeal

[2]PFS is the abbreviation for protospacer flanking site.

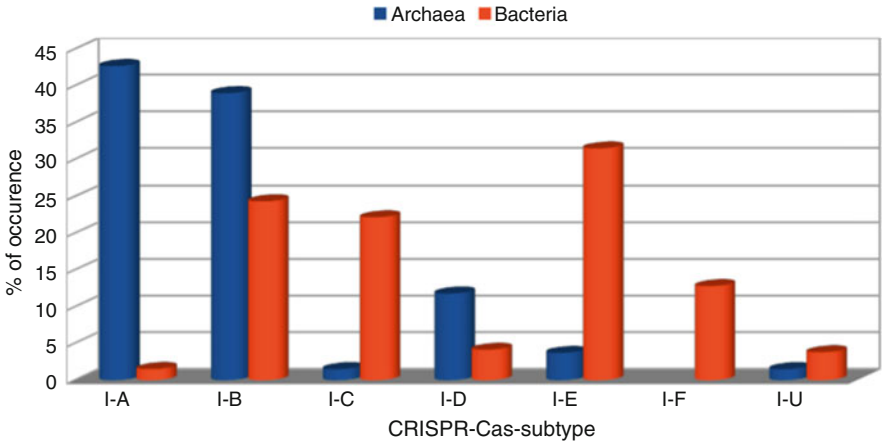

Fig. 11.2 Distribution of type I subtypes in Archaea. Subtypes I-A and I-B are the dominant subtypes in Archaea, subtypes I-C, I-D, I-E and I-U are only present in few archaeal systems, whereas subtype I-F systems are completely absent. In Bacteria however, all subtypes are present with subtypes I-B, I-C and I-E predominant

genomes (Burstein et al. 2017; Makarova et al. 2015; Mohanraju et al. 2016): Cas9 protein genes were identified in a metagenomic analysis in two uncultured nanoarchaeal genomes: *Candidatus* Micrarchaeum acidiphilum ARMAN-1 and *Candidatus* Parvarchaeum acidiphilum ARMAN-4 (Burstein et al. 2017) and a putative type V system could be identified in *Candidatus* Methanomethylophilus alvus (Makarova et al. 2015).

Archaea encode more type III systems than bacteria but the most prevalent systems in archaea are the type I systems with type I-A and I-B being the most abundant (Fig. 11.2). Interestingly, type I-F is completely absent from archaeal genomes (Fig. 11.2) (Makarova et al. 2015; Staals and Brouns 2013). The distribution of CRISPR-Cas types in the archaeal domain is not uniform: crenarchaeota encode mostly type I-A and III-B systems, whereas in euryarchaeota a greater diversity is found with examples of type I-A, -B, -D and type III-A (Makarova et al. 2015; Vestergaard et al. 2014). The type I-B systems are overrepresented within the Euryarchaeota and are most abundant in Haloarchaea (Makarova et al. 2015; Vestergaard et al. 2014). Haloarchaea thrive in the most saline habitats found on earth, e.g., salterns, salt lakes, tidal evaporation ponds, deep-sea salt domes, salt mines, salty soils and anthropogenic salt-dominated environments, such as salted-fish-fermented foods (Oren 2006). They can tolerate salt concentrations up to saturation but also depend on a species-specific minimal salinity within their environment (Oren 2006).

Regarding their CRISPR-Cas content, haloarchaea also form a coherent group; 12 of the 24 publicly available genomes possess complete CRISPR-Cas systems, all of subtype I-B (Fig. 11.3). Most haloarchaea contain only one CRISPR-Cas subtype, and only two strains (*Haloquadratum walsbyi* C23 and *Halorubrum*

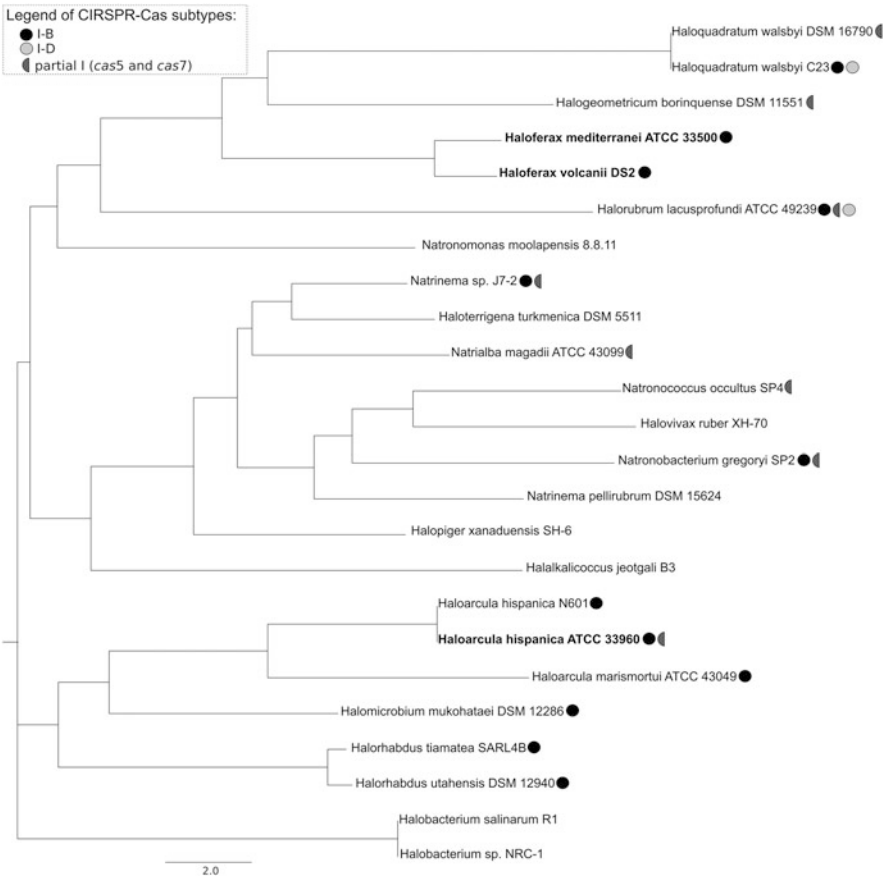

Fig. 11.3 Phylogenetic distribution of CRISPR-Cas systems in Haloarchaea. This phylogenetic tree was constructed using all of the haloarchaeal genomes that are available in public databases (as of October 2016). The presence and type of CRISPR-Cas system found in each species are given. The distribution of *cas* gene cassettes does not show a pattern that correlates with the phylogenetic relationship of the haloarchaea depicted. All of the CRISPR-Cas-positive species exclusively encode type I systems, and most of them only possess a single *cas* gene cassette of subtype I-B (*black dot*). Dual CRISPR-Cas systems are rare and represent a combination of subtypes I-B (*black dot*) and I-D (*grey dot*). Moreover, partial *cas* gene cassettes comprising only cas5 and cas7 genes (*grey halfmoon*) are the sole trace of a CRISPR-Cas system in some species but also co-occur with complete systems of both subtypes. The strains referred to in the text are given in bold

lacusprofundi ATCC49239) encode two different complete CRISPR-Cas systems of subtypes I-B and I-D. Interestingly, some haloarchaeal genomes contain isolated *cas* genes that represent partial effector modules comprising *cas* genes *cas5* and *cas7*. These partial *cas* gene clusters are found in isolation in four species (*Haloquadratum walsbyi* DSM16790, *Halogeometricum borinquense* DSM11551, *Natrialba magadii* ATCC43099, and *Natronococcus occultus* SP4). In *Natrinema*

sp. J7-2, *Natronobacterium gregoryi* SP2 and *Haloarcula hispanica* ATCC 33960, orphan *cas* genes accompany a subtype I-B *cas* gene cluster, whereas in *Hrr*[3]. *lacusprofundi* ATCC49239, a partial cluster is present together with type I-B and I-D systems. Five representatives of haloarchaea were completely devoid of both CRISPR loci and *cas* genes, whereas four strains (*Halogeometricum borinquense* DSM 11551, *Haloquadratum walsbyi* DSM 16790, *Haloterrigena turkmenica* DSM 5511, *Natrinema pellirubrum* DSM 15624) were missing *cas* genes but possessed CRISPR loci, so called orphan CRISPR loci.

The presence and absence, as well as the distribution of CRISPR-Cas types and combinations of subtypes, do not reflect the phylogenetic relationships among species. This is not only true in haloarchaea but was observed throughout all classification efforts (Garrett et al. 2011; Haft et al. 2005; Makarova et al. 2011, 2015; Mohanraju et al. 2016; Vestergaard et al. 2014). An uneven distribution pattern reflects the dynamic nature and rapid pace of evolution of the CRISPR-Cas components as well as the system's propensity to be transmitted by mobile genetic elements or to be lost due to self-targeting or selective pressure, favouring the uptake of mobile DNA elements (Makarova et al. 2015; Shah and Garrett 2011). Thus, the distribution of CRISPR-Cas activity may reflect the balance between the costs and benefits of maintaining a CRISPR-Cas system in the individual habitat of the respective species (Jiang et al. 2013).

A growing body of information on haloarchaeal CRISPR-Cas systems is available and has, until now, been concentrated on subtype I-B and focussed on *Haloferax volcanii* (Brendel et al. 2014; Cass et al. 2015; Fischer et al. 2012; Maier et al. 2012, 2013a, b, 2015b; Stachler and Marchfelder 2016; Stoll et al. 2013), *Haloferax mediterranei* (Li et al. 2013) and *Haloarcula hispanica* (Li et al. 2014a, b; Wang et al. 2016). The signature gene of this subtype is *cas8b*, and the *cas* gene clusters show a conserved arrangement (Makarova et al. 2015), whereas the number and location of the associated CRISPR loci differ widely (Fig. 11.4). However, the nature of the repeats within these CRISPR loci is strictly conserved within the Haloarchaea, resulting in a near-identical repeat sequence and length over the phylogenetic tree (Fig. 11.5). Aspects of all stages of CRISPR-Cas activity have been studied in different haloarchaeal systems, and the resulting picture is summarized in the following paragraphs.

11.3 The Adaptation Process in *Haloarcula hispanica*

The most striking feature of the CRISPR-Cas defence is its capacity to adapt to previously unknown invaders (as reviewed in (Amitai and Sorek 2016; Sternberg et al. 2016)). Until a decade ago, adaptive immunity was exclusively assigned to eukaryotic organisms, a paradigm that was swept aside by the characterization of

[3] *Halorubrum*.

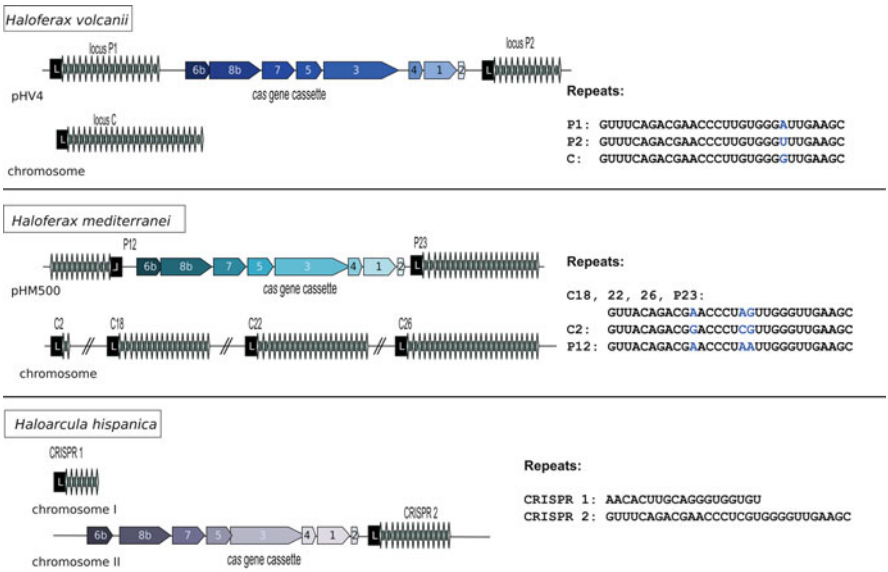

Fig. 11.4 The CRISPR-Cas type I-B systems of the haloarchaeal species discussed in this review. All three species possess a *cas* gene cassette of subtype I-B comprising eight *cas* genes. Gene synteny is conserved, whereas intergenic spacing as well as individual gene sequences are not. A characteristic of type I systems is the presence of the nuclease Cas3, whereas the Cas8b protein further characterizes a subtype I-B system. The *cas* genes are accompanied by a species-specific number of associated CRISPR loci. The repeat sequences within each locus are the same, whereas the sequences of different loci found within one genome vary in few positions (bold, *blue*). In *Haloferax volcanii*, the *cas* gene cassette is encoded on the pHV4 plasmid and is flanked by two CRISPR loci. A third locus is located on the primary chromosome. Laboratory strain H119 shows a deletion of 23 spacers within locus P1 with respect to the genome sequence published for the type strain *Hfx. volcanii* DS2 (Fischer et al. 2012). The repeat sequences of each locus are identical except for position 23. *Haloferax mediterranei* possesses six CRISPR loci. There are two loci flanking the *cas* gene cassette on plasmid pHM500 and four loci distributed on the primary chromosome. The repeat sequences in the four loci are identical, whereas loci C2 and P12 deviate at two positions (bold, *blue*). The *Haloarcula hispanica* CRISPR-Cas system is found on chromosome II. Downstream of the *cas* gene cassette is a single CRISPR locus (CRISPR2). CRISPR2 is also the only locus that encodes a full-length repeat sequence (with respect to the conserved repeat sequence given in Fig. 11.5). CRISPR1 on chromosome I comprises only repeats of 19 nucleotides, which widely deviate from the conserved haloarchaeal repeat (Fig. 11.5). Whether this locus actually results in mature crRNAs is unknown

CRISPR-Cas systems. Early on, the spacer content of CRISPR loci was linked to environmental sequences, such as phages or transposable elements (Bolotin et al. 2005; Mojica et al. 2005; Pourcel et al. 2005), conferring adaptive phage resistance (Barrangou et al. 2007; Brouns et al. 2008; Hale et al. 2009; Marraffini and Sontheimer 2008). The adaptation step expands the CRISPR locus by one repeat-spacer unit, and new spacers are preferentially added to the leader end of the locus, resulting in a near-chronological record of past encounters. First, the intruding nucleic acid has to be identified, and a small portion of its sequence known as the

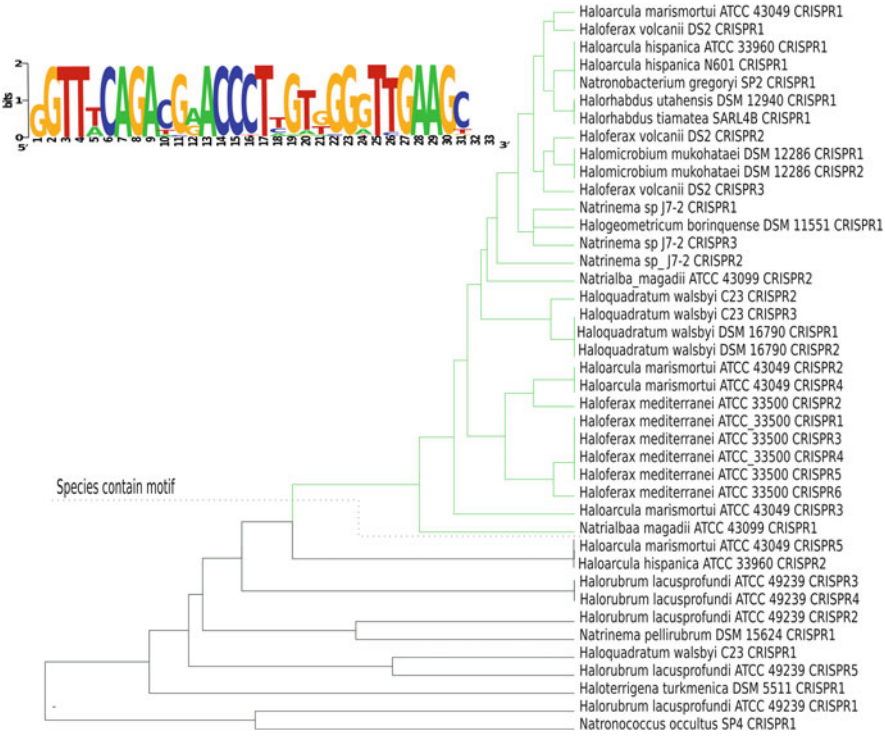

Fig. 11.5 The haloarchaeal repeat sequence is highly conserved and nearly identical throughout Haloarchaea. The repeat sequences found in the haloarchaeal-encoded CRISPR loci (listed on the *right*) have been combined into a sequence logo using the software WebLogo (Crooks et al. 2004), and the corresponding phylogenetic tree with repeat conservation is provided on the *right* (Lange et al. 2013; Alkhnbashi et al. 2014). The overall sequence conservation of the haloarchaeal repeat is very high, and only a few positions show interspecies differences. This trend, together with the conservation of the processing site utilized by Cas6 endonuclease during crRNA maturation, gives rise to highly uniform crRNA populations

protospacer has to be selected for integration. Crucial for this step in type I systems is a conserved series of a few nucleotides upstream of the protospacer, which is the PAM (Deveau et al. 2008; Erdmann and Garrett 2012; Mojica et al. 2009; Shmakov et al. 2015; Swarts et al. 2012; Wang et al. 2015). The PAM not only allows for the selection of a sequence as a protospacer during adaptation but also ensures discrimination between endogenous CRISPR loci that encode the spacer and the invader carrying the protospacer during the interference stage (Shah et al. 2013). The Cas proteins that are essential for the adaptation step are Cas1 and Cas2, which form a complex (Nunez et al. 2015; Plagens et al. 2012; Yosef et al. 2012). The concerted activity of the Cas1/Cas2 complex leads to the integration of the new spacer at the leader-repeat junction via a transposase/integrase-like mechanism, as shown in studies of the *E. coli* I-E system (Arslan et al. 2014; Nunez et al. 2015; Yosef et al. 2012). In *E. coli* (type I-E) DNA polymerase I and presumably other not

yet identifies factors are involved (Ivancic-Bace et al. 2015) and adaptation depends on RecBCD activity occurring at sites of double strand breaks found e.g. at replication forks (Ivancic-Bace et al. 2015; Levy et al. 2015). In type I-B systems the Cas4 protein is also required for adaptation but details on its involvement are not unravelled yet.

In addition to this *de novo* capture of spacers known as naïve adaptation, a second form of spacer acquisition called primed adaptation was shown (Datsenko et al. 2012; Fineran et al. 2014; Künne et al. 2016; Li et al. 2014b; Richter et al. 2014; Semenova et al. 2016; Swarts et al. 2012; Vorontsova et al. 2015). During primed adaptation, a pre-existing spacer induces a positive-feedback loop, which leads to enhanced spacer acquisition from the targeted genetic element. A non-perfect match between a pre-existing crRNA and an invader DNA results in a defective interference reaction, and the elimination of the invader is not achieved; however, enhanced acquisition activity is induced (Datsenko et al. 2012; Fineran et al. 2014). Accordingly, in addition to the key acquisition proteins Cas1 and Cas2, primed adaptation also requires the presence of the Cascade interference complex as well as the Cas3 nuclease (Datsenko et al. 2012; Künne et al. 2016; Li et al. 2014b; Swarts et al. 2012). The co-occurrence of naïve and primed adaptation has so far been demonstrated in the *E. coli* type I-E system (Datsenko et al. 2012; Fineran et al. 2014; Swarts et al. 2012; Yosef et al. 2012), and the type I-F systems of *Pectobacterium atrosepticum* and *Pseudomonas aeruginosa* (subtype I-F, (Richter et al. 2014; Staals et al. 2016; Vorontsova et al. 2015)), whereas in the *Har. hispanica* subtype I-B system, only primed adaptation seems to be employed (Li et al. 2014b). *Har. hispanica* is the only haloarchaeal system for which adaptation could be shown to date. The deletion of the adaptation genes *cas1*, *cas2* and *cas4*, as well as the deletion of the interference module (*cas5–8*) in its entirety or the effector nuclease Cas3, will render *Har. hispanica* cells incapable of acquiring new spacers (Li et al. 2014b). In accordance with the necessity of the interference machinery, this process is strictly limited to primed adaptation, as the deletion of the priming spacer, with limited complementarity to the invader sequence, likewise hinders the integration of new spacers (Li et al. 2014b). During the priming process, a Cascade loaded with the imperfectly matched crRNA binds to the protospacer region of the invader in low-fidelity binding mode, triggering the priming process, as shown by FRET (Förster resonance energy transfer) analysis in *E. coli* (Blosser et al. 2015). Cas3 was speculated to be involved in the provision of acquisition substrates (Ivancic-Bace et al. 2015; Swarts et al. 2012). This speculation was supported by the observation that mutation of conserved residues within *Har. hispanica* Cas3 clearly show the active involvement of both the HD nuclease and the DxD/H-helicase domain in spacer acquisition (Li et al. 2014b). Recent work in *E. coli* further confirmed it: Cas3 degradation products were bound by the Cas1-Cas2 complex and integrated as new spacers (Künne et al. 2016). In *P. atrosepticum* (I-F system) interference promotes a targeted spacer acquisition process similar to priming (Staals et al. 2016).

Primed adaptation results in a biased sampling of new spacers with respect to the location of the priming protospacer. In *E. coli*, the DNA strand from which the new

spacers are derived from is the same as that of the priming protospacer (Datsenko et al. 2012; Fineran et al. 2014). However, in *Har. hispanica*, the pattern is more diverse. Upstream of the imperfect match, spacers stem from the non-target strand, and their orientation matches the priming protospacer, whereas downstream, the target strand is the preferred source of spacers with the opposite directionality (Li et al. 2014b). A similar distorted acquisition pattern from both strands has also been observed in type I-F (Richter et al. 2014). The spacers acquired by *Har. hispanica* over the course of these first experiments were all sampled exclusively from the infecting viral particles, and the cognate protospacers were always preceded by the PAM TTC (Li et al. 2014b). Li and co-workers further used a mutational approach, presenting *Har. hispanica* cells with all possible three-nucleotide PAM combinations within an invader plasmid to study the motif's impact on the initiation of primed adaptation as well as its variability (Li et al. 2014a). Twenty-three of the 64 tested PAM sequences induced primed adaptation when the 5′ end of the protospacer was targeted by the priming spacer, these PAMs were termed priming permissive (Fig. 11.6) (Li et al. 2014a). Further investigation

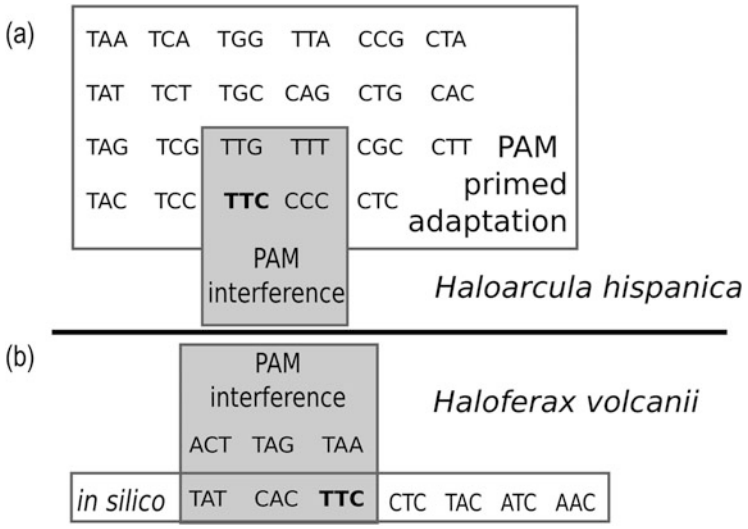

Fig. 11.6 Overview of PAM requirements during primed adaptation and interference. (**a**) *Haloarcula hispanica* is the first haloarchaeon for which primed adaptation was studied. 23 out of 64 possible PAM sequences triggered primed adaptation (Li et al. 2014a, b). In contrast, interference was triggered only with four PAM sequences (highlighted in *grey*) (Li et al. 2014a). (**b**) For *Hfx. volcanii*, the PAM requirements have only been determined for the interference stage. Here, six sequences trigger successful elimination of an invader DNA (Fischer et al. 2012). Additional *in silico* analyses revealed seven motifs found upstream of sequences matching spacers of the *Hfx. volcanii* CRISPR loci (Maier et al. 2015a). Three of them are identical to the identified interference PAM sequences. Despite their close phylogenetic relationship, both haloarchaea only share one PAM motif: TTC (marked in bold)

also revealed that these sequences are not sensed by a base-pairing mechanism but rather through the authentication of the PAM sequence (Li et al. 2014a). In addition, repeat sequences flanking the targeted protospacer do not impair priming as long as a cognate priming-permissive PAM is present (Li et al. 2014a). Similar findings were made in *E. coli*, here some PAM sequences can trigger both, interference and priming, but more PAM sequences were permissive for primed adaptation than for direct interference (Fineran et al. 2014; Li et al. 2014a).

Requirements for the integration of new spacers were also studied in detail in *Har. hispanica* (Wang et al. 2016). The sequences surrounding the leader-repeat junction are highly conserved within Haloarchaea and the conserved leader sequence plays a critical role during spacer integration (Wang et al. 2016). The leader-proximal cut occurs consistently at the leader-repeat junction (Fig. 11.7, integration site 1). However, the leader distal cut site does not have specific sequence requirements but is located at a constant distance to the second conserved repeat motif GTGGG (Fig. 11.7, integration site 2). A mutational analysis of the repeat sequence revealed that two conserved motifs in the repeat sequence are required for integration of new spacers (Fig. 11.7). The first motif (AACCC) needs to be 10 base pairs downstream of the leader-repeat junction and presumably serves as docking site for the integrase complex. The second motif (GTGGG) seems to be the anchor for a molecular ruler to direct the second cut 10 base pairs downstream thereby determining the size of the repeat duplication. Analysis of adaptation in *E. coli* confirms the presence of a ruler mechanism to define repeat length (Goren et al. 2016). Here, two rulers are employed, both anchored in the repeat sequence. Whether this model on the governing of the spacing of integration events is also true for other systems has yet to be determined. Moreover, an analysis of other haloarchaeal species will reveal whether adaptation in Haloarchaea is truly limited to a priming process.

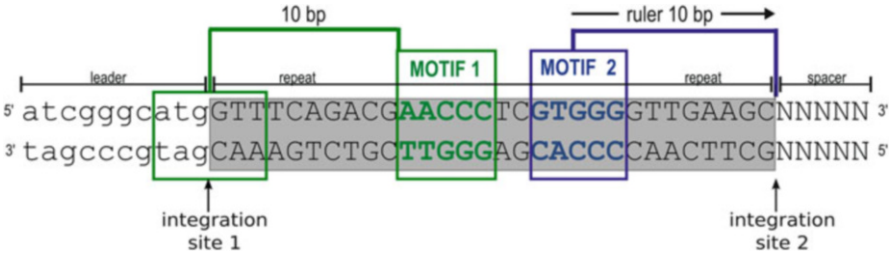

Fig. 11.7 Primed adaptation in *Har. hispanica:* motifs governing the integration process. A mutational analysis revealed sequence-specific recognition of the sequence spanning the leader-repeat-junction as well as two important motifs in the middle of the repeat sequence. The first motif (motif 1 depicted in *green*) has to be located ten nucleotides downstream of the leader-repeat-junction, whereas motif 2 (shown in *blue*) serves as an anchor-point to direct the second cleavage

11.4 Expression of the crRNA and Assembly of the Cascade Complex

The key player in CRISPR-Cas interference is the crRNA. The small guide that confers invader specificity to the Cascade complex is allocated during the expression stage of CRISPR-Cas activity. The crRNA is bound to and positioned within the Cascade by a subset of Cas proteins. Each complex is loaded with an individual guide that allows for the sequence-specific identification of an invader, equipping the cell with a multitude of effector complexes (Brouns et al. 2008; Jore et al. 2011; Künne et al. 2014).

The transcription of CRISPR loci is driven by a promoter region within the leader sequence (Pul et al. 2010). In *Hfx. volcanii* as well as in *Hfx. mediterranei*, the expression of the long precursor, pre-crRNA, is constitutive (Fischer et al. 2012; Li et al. 2013). The release of the crRNAs follows through the processing of the repeat region, catalysed by the Cas6b protein. This was confirmed in *Hfx. volcanii* and *Hfx. mediterranei*, where crRNA production is lost upon deletion of the *cas6b* gene (Brendel et al. 2014; Li et al. 2013). Apart from Cas6b, multiple other Cas proteins are involved in the maintenance of a stable crRNA population within the cell. In *Hfx. volcanii*, a deletion of the *cas5* or *cas7* gene does not impair but rather severely lessens the steady-state level of crRNA, which indicates that there is a Cas5- and Cas7-mediated protection against degradation (Brendel et al. 2014). The protective effect of Cas5 and Cas7 is even more pronounced in the *Hfx. mediterranei* system, in their absence, no mature crRNA is detectable (Li et al. 2013).

By contrast to *cas6b* deletion, the loss of *cas5/7* clearly leaves endonucleolytic processing intact because a leader-first repeat product still accumulates in *Hfx. mediterranei* (Li et al. 2013). As revealed by structural studies with the subtype I-E Cascade in *E. coli*, Cas5 binds to the repeat-derived 5′-handle, whereas Cas7 covers the spacer sequence, thereby enclosing the crRNA within the Cascade and making it less accessible for the degradation machinery (Jackson et al. 2014; Mulepati et al. 2014; Zhao et al. 2014). Both Cas proteins are also integral parts of the *Hfx. volcanii* Cascade, which in addition includes Cas6b (Fig. 11.8) (Brendel et al. 2014). The Cas8b protein was only occasionally obtained and therefore seems to be only loosely associated with the *Haloferax* Cascade complex. This finding is also mirrored by the minor stabilizing effect of Cas8b on the crRNA population in both *Haloferax* species (Brendel et al. 2014; Li et al. 2013).

The crRNA itself also affects the structure and composition of the effector complex. The Cascade complex of the type I-B system in *Haloferax* has not yet been structurally characterized, but a combination of mass spectrometry and intensity-based absolute quantification (iBAQ) identified the core complex as being composed of Cas5, Cas6b and Cas7 in a ratio of 1.7:1:8.5 (Brendel et al. 2014). This complex differs from the composition of the *E. coli* type I-E Cascade, for which the stoichiometry was also determined as Cas5, Cas6, Cas7, Cas8, and Cse2: 1:1:6:1:2 (Jore et al. 2011; Wiedenheft et al. 2011). The small subunit Cse2

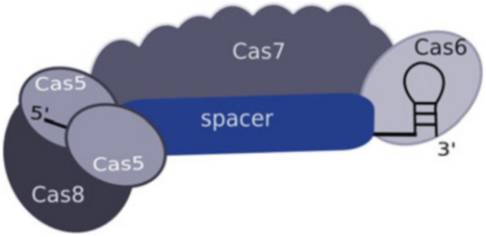

Fig. 11.8 Potential structures of Cascade complexes in *Hfx. volcanii*. Since structural data are not available for the *Haloferax* Cascade or any I-B Cascade complex, a schematic representation based on the published structure of the *E. coli* type I-E complex is given. Co-purification approaches combined with quantitative mass spectrometry identified a Cascade complex composed of Cas5, Cas6, Cas7 in a stoichiometry of 1.7:1:8.5 (Brendel et al. 2014). Cas8b seems to be loosely associated and could only be occasionally co-purified (Brendel et al. 2014). A minimal stable complex might be formed that includes only Cas5 and the Cas7 backbone alongside the crRNA, as Cas6b is dispensable for interference (Brendel et al. 2014; Maier et al. 2015b). The minimal crRNA that elicits an interference reaction only comprises the 5′-handle and the spacer sequence; thus, the minimal Cascade may well be further reduced by omitting the crRNA 3′ handle (Maier et al. 2015b)

and the Cas8 protein are integral parts of the *E. coli* I-E Cascade, which is the most striking difference, but the composition of the core *Hfx. volcanii* I-B Cascade also shows two additional copies of Cas7 (Brendel et al. 2014). This composition might reflect differences in the length of the crRNA because the spacer length in *Haloferax* is 34–39 nt, as opposed to the 32 nt in *E. coli*. Given that Cas7 forms the backbone of Cascade receiving the spacer portion of the crRNA, additional subunits are needed to cover the entire 2- to 7-nt-longer spacer sequence found in *Haloferax*. The elongation of the Cascade backbone to accommodate an elongated crRNA has also been observed in a study that analysed the subunit composition of *Shewanella putrefaciens* I-F Cascade. Upon the extension of the spacer portion of the crRNA, more Cas7 subunits are incorporated (Gleditzsch et al. 2016).

Further analysis of *cas* gene deletion mutants in *Hfx. mediterranei* revealed a negative effect of Cas1, 3 and 4 on the crRNA level (Li et al. 2013), however, it is not clear how these proteins might contribute to crRNA stabilization or influence pre-crRNA expression.

11.5 The crRNA Populations of Haloarchaeal CRISPR-Cas Systems

The position of the Cas6 cleavage site within the repeat regions of the pre-crRNA is highly conserved in type I systems and shows the tight evolutionary and phylogenetic link between the repeat sequence and Cas6 protein family (Kunin et al. 2007; Wang et al. 2012). However, individual Cas6 proteins share neither a common mode for substrate binding nor conserved catalytic residues (Brendel et al. 2014).

Their amino acid sequences show limited conservation with only two common motifs: the ferredoxin fold and a glycine-rich motif (Li 2015). The crRNAs of type I systems are consistently composed of the spacer sequence accompanied by an eight-nucleotide 5′-handle and the remainder of the downstream repeat as a 3′-handle (Charpentier et al. 2015). The extraordinary conservation of the repeat sequences present in haloarchaeal CRISPR-Cas systems results in near-identical crRNA flanking sequences, and the conserved repeat length results in 22 nt-long 3′-handles in almost all haloarchaeal species.

Interestingly, the three CRISPR loci of *Hfx. volcanii* each differ in their repeat sequence at position 23, resulting in a mixed population of mature crRNAs starting with either U, A or G as the first nucleotide of the 5′-handle (Fig. 11.4) (Fischer et al. 2012). The genome of *Hfx. mediterranei* encodes six CRISPR loci, also sharing a common repeat sequence with slight variations in the repeat of locus C2 at positions 11 (A to G) and 17 (A to C), and the repeat sequence of locus P12 deviates at position 18 (G to A) (Fig. 11.4). Those variants also result in a population of crRNAs with varying 3′-handle sequences (Li et al. 2013).

Despite the close phylogenetic relationship of both *Haloferax* species, they differ in the size distribution of the crRNA population that is detectable *in vivo* (Fig. 11.9). The analysis of crRNA sequences in *Hfx. mediterranei* by CR-RT-PCR (circularized-RNA RT-PCR) revealed one population of mature crRNAs with a size range from 64 to 68 nt and with differences accounted for by the varying spacer lengths (approximately 34–39 nt, as expected) (Li et al. 2013). However, an RNA-Seq approach in *Hfx. volcanii* identified a second group of crRNAs that is stably maintained separate from the dominant crRNA population of 64–69 nt due to the spacer length (Maier et al. 2015b). These crRNAs are substantially shorter due to having a 3′-handle of only five nucleotides. A similar trimming of mature crRNAs has been reported for other type I-B systems of *Clostridium thermocellum* and *Methanococcus maripaludis* and is assumed to be characteristic of type I-B as well as I-A and I-D systems in contrast to the type I-C, I-E and I-F systems featuring non-trimmed crRNAs (Charpentier et al. 2015). This variety illustrates the diversity of CRISPR-Cas mechanisms, reaching beyond the subtype-level and making it even more important to study a wide variety of CRISPR-Cas systems in different species to complete the picture of this most elaborate defence system.

Moreover, the crRNA population of *Hfx. volcanii* revealed that crRNAs originating from the same CRISPR locus are not present in equal amounts (Maier et al. 2013a), an observation confirmed in several other organisms of different subtypes (Deng et al. 2012; Hale et al. 2012; Nickel et al. 2013; Richter et al. 2012; Scholz et al. 2013; Zhang et al. 2012). This finding might reflect an imminent technical problem in the currently available RNA-Seq approach but it might also be of biological relevance. Furthermore, as shown for *Hfx. volcanii*, the different crRNAs diverged in their ability to fend off the plasmid invader (Maier et al. 2013a). The different crRNAs present in the cell vary in their spacer sequence, which may not only contain signals that trigger a faster degradation of some of the molecules but may also influence the effectivity of Cascade binding. In addition, they may

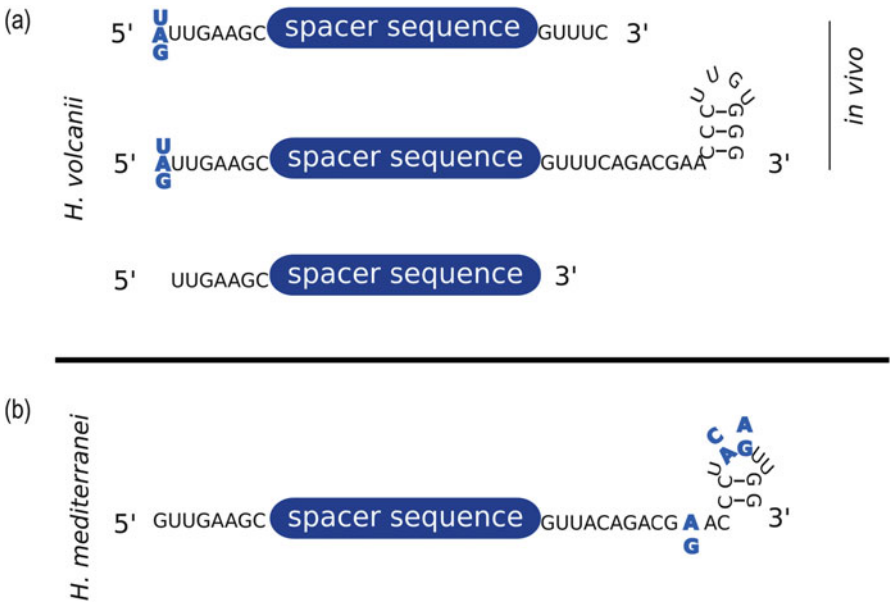

Fig. 11.9 Different crRNA molecules *in Hfx. volcanii* and *Hfx. mediterranei*. (**a**) In addition to the spacer sequence, the long form of the crRNA identified in vivo in both *Haloferax* strains possesses an 8-nucleotide 5′ handle and a 22 nucleotide long 3′ handle (Li et al. 2013; Maier et al. 2013a). In *Hfx. volcanii,* the first nucleotide of the 5′ handle differs due to sequence variation within the three CRISPR loci. The varying repeat sequences in *Hfx. mediterranei* result in a mixed population of crRNAs with variable 3′ handles. The RNA-Seq analysis of the crRNA pool in *Hfx. volcanii* also revealed a shortened crRNA variant with only five nucleotides as the 3′ handle and seven instead of eight nucleotides at the 5′ handle (Maier et al. 2015b). Moreover, mutational analysis demonstrated that the crRNA is still active when the 3′ handle is completely removed (Maier et al. 2015b)

influence the microarchitecture and topology of the Cascade complex and thereby the efficiency of the interaction with the target.

The crRNA structure has so far only been investigated in *Hfx. volcanii*, and although the 22-nucleotide 3′-handle encodes a set of inverted repeats, offering the possibility of forming a hair-pin structure at the very 3′ end, no such structure has been detected in *in vitro* studies (Fischer et al. 2012). Nevertheless, a hairpin structure might be stabilized upon interaction with the Cas6b protein during processing, as seen in *Thermus thermophilus* and *Sulfolobus solfataricus* (Niewoehner et al. 2014; Shao and Li 2013).

The characteristics of the crRNA with importance beyond processing were studied in *Hfx. volcanii* using a system for the Cas6-independent generation of crRNAs based on the tRNA-maturation machinery (Maier et al. 2015b). The independent biogenesis pathway results in a crRNA with a 5′-phosphate and 3′-hydroxyl group, in contrast to a crRNA processed by Cas6b possessing a 5′-hydroxyl and 2′-3′-cyclic phosphate group. This independently generated crRNA (termed icrRNA) was active, therefore, neither the loading of a crRNA

into Cascade nor the interference reaction depends on the chemical nature of the end groups (Maier et al. 2015b). A mutational analysis of independently generated crRNAs revealed that the 3′-handle of the crRNA was completely dispensable for the *in vivo* interference reaction (Maier et al. 2015b). A comparison with the structure of the type I Cascade complexes of *E. coli* showed that the 3′-handle would come into contact with Cas6b (Jackson et al. 2014; Mulepati et al. 2014; Zhao et al. 2014). However, in a *Hfx. volcanii* strain capable of Cas6b-independent crRNA maturation, *cas6b* could be deleted without affecting the interference step (Maier et al. 2015b). This finding implies that Cas6b is not an essential part of the type I-B Cascade complex for the interference step (Fig. 11.8). The crRNA 5′-handle, by contrast, is indispensable for crRNA function, only the first nucleotide can be removed without loss of activity (Maier et al. 2015b). Inference from the atomic structures of type I-E Cascade showed that in *E. coli* is bound by the Cas5 subunit (Jackson et al. 2014; Mulepati et al. 2014; Zhao et al. 2014), which was also shown to be an integral part of the *Haloferax* type I-B Cascade (Brendel et al. 2014).

11.6 Determinants for a Successful Defence Reaction in Subtype I-B

CRISPR-Cas systems have recently attracted attention as a molecular biological tool that out-competes all the available nucleic-acid targeting proteins because its targeting activity is based on an easily interchangeable module: crRNA. Through the embodied spacer sequence, different crRNAs guide the Cascade complex to a defined targeting site within an invading nucleic acid. This identification depends on base pairing between the spacer part of the crRNA and the target molecule (Künne et al. 2014). A mutational analysis of the protospacer sequence within a plasmid invader was used to determine how strictly defined this interaction is regarding the *Hfx. volcanii* type I-B system (Maier et al. 2013a). The first ten nucleotides of the spacer sequence were identified as being critical for triggering a defence reaction. Within this sequence, which is denoted as the seed region, only a mismatch at position 6 is tolerated (Fig. 11.10). A similar seed sequence was also determined for *E. coli* and *P. aeruginosa* (Künne et al. 2014; Semenova et al. 2011; Wiedenheft et al. 2011). For the *E. coli* system, every sixth position within the crRNA is not involved in a base-pairing activity (Semenova et al. 2011). As structural data show, the thumb domain of the Cas7 proteins in the *E. coli* Cascade extrude every sixth nucleotide from the axis that runs down the Cas7 backbone, rendering it inaccessible for base pairing with the protospacer region (Jackson et al. 2014; Mulepati et al. 2014; Zhao et al. 2014). Increments of six could not be confirmed in the *Hfx. volcanii* analysis, but this difference might very well reflect the aforementioned differences in the Cascade composition (Maier et al. 2013a).

An important prerequisite for having and keeping a CRISPR-Cas defence system is, that self-targeting is excluded, since this can be fatal for the cells. Such an

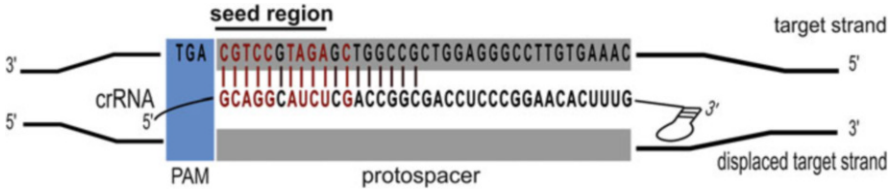

Fig. 11.10 A seed sequence is required for interference in *Hfx. volcanii*. During invader recognition, the crRNA base pairs with the protospacer region of the invading DNA. Base pairing over a ten-nucleotide-long non-contiguous seed sequence elicits the interference reaction. Essential base pairs are shown in *red*. Pairing at position six is not required, but *Hfx. volcanii* does not exhibit a six nucleotide increment as seen in the spacer-protospacer interactions in *E. coli* (Maier et al. 2013a)

auto-immune reaction is prevented by the absence of the PAM in the host DNA. The role of PAM sequences in adaptation has already been discussed, but these sequences also play an important role during the interference stage (Deveau et al. 2008; Mojica et al. 2009). PAM sequences can be found in type I, type II and type V CRISPR-Cas systems and are unique features of the protospacer. Despite perfect base pairing within the seed sequence, interference takes place only if a cognate PAM is present at the 5′ end of the protospacer sequence in type I and type V systems and at the 3′ end in type II systems (Shah et al. 2013; Zetsche et al. 2015). PAM sequences of haloarchaeal species could not be directly inferred by comparing the spacer contents and publicly available sequences of mobile genetic elements because haloarchaeal viruses are grossly underrepresented in public databases (Fischer et al. 2012). Moreover, the population of mobile genetic elements present today likely differs substantially from the one that was present on the isolation date of the laboratory strains under investigation. Therefore, the PAM sequences of the *Hfx. volcanii* type I-B system have been identified *in vivo* using a mutational approach based on a plasmid invader (Fischer et al. 2012). A systematic analysis of all possible three nucleotide sequences preceding a protospacer revealed the following six PAM sequences: ACT, CAC, TTC, TAT, TAG, and TAA (Fig. 11.6).

The stimulation of an interference reaction by more than one sequence motif is a strategy for coping with the divergence of invader populations, rendering escape via individual PAM mutations less likely. Moreover, this stimulation increases the possibility that closely related foreign elements are also susceptible to CRISPR-Cas interference. The authentication of the PAM sequence is a crucial step in the transition of Cascade from a DNA-sensing to a DNA-degrading complex. Studies of *E. coli* type I-E Cascade show that upon detection of a cognate PAM sequence, the conformation of the Cascade is changed and the processing endonuclease Cas3 is recruited (Hochstrasser et al. 2014). The subunit responsible for determining the PAM identity in type I-E systems is the large subunit Cas8e (Sashital et al. 2012). Accordingly, upon the deletion of Cas8b in *Haloferax*, the interference was lost without affecting the crRNA level or stability (Cass et al. 2015). More importantly, the response of Cas8b variants with mutated conserved residues varied with regard to the PAM sequence presented by the invader (Cass et al. 2015). However, the

exact mechanism for the read-out of PAM sequences is still under investigation. Although *Har. hispanica* shares the same subtype and a near-identical repeat sequence with *Hfx. volcanii*, the *Har. hispanica* CRISPR-Cas system responds to only four PAM sequences, namely TTT, TTC, TTG, and CCC, and besides TTC, no other PAM is shared between both organisms (Fig. 11.6) (Li et al. 2014a). The Cas8b proteins found in both species only share 22.6% sequence identity, and given the likely role of Cas8b as the PAM-sensing Cascade subunit, this low similarity might account for the low conservation of PAM sequences (Li et al. 2014a). Moreover, findings from a bioinformatics analysis of the spacer content of *Hqr*[4]. *walsbyi* provide support that a certain degree of PAM sequence conservation is present (Fischer et al. 2012; Garcia-Heredia et al. 2012). Several matches to viral contigs from the metavirome data of the isolation sites reveal protospacers preceded by the PAM TTC. A recent bioinformatics analysis of PAM sequences in *Hfx. volcanii* could match eight of the *Hfx. volcanii* spacers to sequences in the database (Maier et al. 2015a). These target sequences are flanked by seven different PAM sequences: TAT, CAC, CTC, TTC, TAC, ATC and AAC at the protospacer 5′-end. Three of them are identical to the experimentally determined PAMs: CAC, TTC and TAT. However, the motifs inferred from *in silico* analysis were obtained by comparing sequences that were not necessarily derived from the same biological context.

Together with the adaptation analysis in *Har. hispanica* (see paragraph above), this evidence illustrates that the requirements for PAM sequences during the adaptation and interference stages are not identical but can overlap. As this trend is also seen in other systems, PAMs have been subdivided into motifs important for adaptation, termed spacer acquisition motif (SAMs) and motifs essential for interference (target interference motif: TIMs) (Shah et al. 2013). These processes rely on different protein machineries: the Cas1/Cas2 complex is interacting with the SAM during naïve adaptation, while the Cascade complex is interacting with TIM during interference. Different binding partners or different conformations of the binding subunit within the complex might result in different PAM demands (Shah et al. 2013).

An analysis of the *Hfx. volcanii* interference reaction revealed an interesting detail that influenced the success of the defence reaction. Whether a plasmid invader effectively triggers an interference reaction depends on the origin of the replication (Maier et al. 2013a). Only the plasmid with a pHV1 origin, replicated by a mechanism depending on an origin recognition complex (ORC), was successfully eliminated (Delmas et al. 2009; Maier et al. 2013a; Norais et al. 2007). By contrast, the type I-B system did not overcome a plasmid based on a pHV2 origin whose replication presumably depended on the Rep protein (Charlebois et al. 1987; Maier et al. 2013a; Woods and Dyall-Smith 1997). The experimental design of the studied plasmids places the targeted protospacer next to the origin of replication. Whether these differences are solely due to steric constraints or reflect a functional interaction requires further analysis.

[4] *Haloquadratum.*

11.7 Using CRISPR-Cas as Tool to Regulate Gene Expression in *Hfx. volcanii*

CRISPR-Cas systems have been developed into a plethora of different tools and have been exploited for numerous scientific analyses outside their natural function in both prokaryotes and eukaryotes (Cong et al. 2013; Fineran and Dy 2014; Sampson and Weiss 2014; Sternberg and Doudna 2015). Given the presence of a cognate PAM sequence, any region of interest can be targeted. The CRISPR-Cas system that is most extensively used in genetic studies is the type II system, which only requires a single effector protein: Cas9. One of the applications is targeted gene regulation, which is performed via a mechanism called CRISPRi (CRISPR interference) (Qi et al. 2013). Here, the expression of a gene specified by the targeting guide RNA is repressed through the binding of a catalytically inactive Cas9 (dCas9) (Qi et al. 2013). The protein is engineered to locate and bind the target sequence as defined by the incorporated crRNA but not to cleave it, thereby preventing or severely hampering transcription initiation or elongation. *Streptococcus pyogenes* dCas9 has successfully been used in eukaryotes as well as in bacteria (Bikard et al. 2013; Qi et al. 2013). The endogenous type I system was likewise repurposed as tool for CRISPRi in *E. coli* by deleting the *cas*3 gene (Luo et al. 2015; Rath et al. 2015). Upon the loss of the targeting endonuclease, Cascade still binds the target region specified by the crRNA but does not cleave it (Luo et al. 2015; Rath et al. 2015).

Molecular biology studies in Archaea become more and more widespread, but tools for transcriptional repression are not available. Due to their extremophilic nature, most archaea pose a challenge to the heterologous expression of proteins, such as Cas9, which is predominantly found in mesophilic bacteria. Therefore, the most straightforward approach is to repurpose the endogenous CRISPR-Cas systems, circumventing the need for the heterologous expression of the Cas9 protein. Similar to the approach used in *E. coli*, the type I-B system in *Hfx. volcanii* has successfully been modified and converted into the first tool for transcriptional repression in archaea (Stachler and Marchfelder 2016). To eliminate the DNA cleavage activity, the *cas*3 gene is deleted and to achieve the efficient downregulation of the targeted gene, the endogenous crRNA population has to be depleted (Stachler and Marchfelder 2016). The latter can be achieved via the deletion of *cas6b* or the deletion of the endogenous CRISPR loci. If Cas6b was no longer present, crRNAs have been provided by the aforementioned icrRNA system (see above) (Maier et al. 2015b; Stachler and Marchfelder 2016). Using this system a plasmid-borne reporter gene, a chromosomal gene, a gene cluster as well as an essential gene were successfully knocked down. The strongest repression effect observed was a down regulation to 8% of the transcript level, highlighting the potential of CRISPRi for archaeal systems (Stachler and Marchfelder 2016).

Various possible targeting regions within the promoter and coding regions of the genes of interest were explored. In general, the crRNAs targeting the promoter region and, more precisely, its template strand are the most efficient, whereas those

directed towards the coding strand or the open reading frame had little or no effect on gene expression (Stachler and Marchfelder 2016). Currently, a clear connection between successful targeting and the characteristics of the crRNA could not be inferred, more experimental data on this topic are required, which might then allow to implement a tool for designing efficient crRNAs.

11.8 Conclusions

Haloarchaea form a coherent group with respect to their CRISPR-Cas content, they all encode type I systems, with the subtype I-B being the most dominant. General characteristics of the archaeal type I-B systems can be drawn on the basis of detailed studies of three haloarchaeal species.

The Cascade-like effector complex analysed in *Hfx. volcanii* closely resembles other type I complexes in terms of its Cas protein composition. The same *cas* genes are consistently indispensable for the *Hfx. mediterranei* CRISPR-Cas activity. *Hfx. volcanii* is the only species that generates two types of crRNAs that differ in the lengths of their 3′ ends. This second processing event has also been described in other subtype I-B systems (Richter et al. 2012) but has not been found in the other haloarchaeal type I-B systems. The shortened crRNA variant of *Hfx. volcanii* lacks part of the 3′-handle responsible for Cas6b binding, probably resulting in a Cascade-like complex lacking this subunit. The minimal requirements for a successful defence reaction in *Hfx. volcanii* support this interpretation because the Cas6b protein has been shown to be dispensable during interference.

The adaptation step was hitherto only investigated in two haloarchaea, namely *Hfx. mediterranei* and *Har. hispanica*. Here, interestingly, only the primed adaptation triggered by the presence of a pre-existing spacer could be shown. The PAM sequences for the adaptation reaction have been determined for *Har. hispanica* revealing 23 PAM sequences that allow primed adaptation. They only partially overlap with PAM sequences, that trigger a defence reaction.

The PAM motifs required for effective interference reactions were systematically analysed in *Hfx. volcanii* revealing that *Hfx. volcanii* responds to six PAM sequences. The *Har. hispanica* and *Hfx. volcanii* PAM requirements overlap in only one motif, illustrating that even closely related haloarchaea differ in their defence requirements. PAM sequences in haloarchaea are situated 5′ to the protospacer sequence on the invading DNA, and similar to other type I systems, interference relies on the presence of a seed sequence as shown in *Hfx. volcanii*.

The information summarized here, along with that of other known type I systems, highlights the subtype-specific and inter-subtype strain-specific peculiarities and differences of type I systems. Further exploration will complete our knowledge on CRISPR-Cas immunity, and subsequently allow the application of CRISPR-Cas systems and their components. The details gathered on CRISPR-Cas immunity in *Hfx. volcanii* enabled the first application of an endogenous archaeal CRISPR-Cas system for targeted gene regulation, meeting a long-standing need for gene regulatory tools in archaea that will further promote archaeal research in many fields.

References

Alkhnbashi OS, Costa F, Shah SA, Garrett RA, Saunders SJ, Backofen R (2014) CRISPRstrand: predicting repeat orientations to determine the crRNA-encoding strand at CRISPR loci. Bioinformatics 30:i489–i496. doi:10.1093/bioinformatics/btu1459

Amitai G, Sorek R (2016) CRISPR-Cas adaptation: insights into the mechanism of action. Nat Rev Microbiol 14:67–76. doi:10.1038/nrmicro.2015.1014. Epub 2016 Jan 11

Arslan Z, Hermanns V, Wurm R, Wagner R, Pul U (2014) Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. Nucleic Acids Res 42:7884–7893. doi:10.1093/nar/gku7510. Epub 2014 Jun 11

Barrangou R (2015) Diversity of CRISPR-Cas immune systems and molecular machines. Genome Biol 16:247. doi:10.1186/s13059-13015-10816-13059

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709–1712

Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res 41:7429–7437. doi: 10.1093/nar/gkt7520. Epub 2013 Jun 12

Blosser TR, Loeff L, Westra ER, Vlot M, Kunne T, Sobota M, Dekker C, Brouns SJ, Joo C (2015) Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. Mol Cell 58:60–70. doi:10.1016/jmolcel201510011028. Epub 2015 Mar 15

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151:2551–2561

Brendel J, Stoll B, Lange SJ, Sharma K, Lenz C, Stachler AE, Maier LK, Richter H, Nickel L, Schmitz RA, Randau L, Allers T, Urlaub H, Backofen R, Marchfelder A (2014) A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crrnas) in Haloferax volcanii. J Biol Chem 289:7164–7177. doi: 10.1074/jbc.M7113.508184. Epub 2014 Jan 23

Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321:960–964

Burstein D, Harrington LB, Strutt SC, Probst AJ, Anantharaman K, Thomas BC, Doudna JA, Banfield JF (2017) New CRISPR-Cas systems from uncultivated microbes. Nature 542:237–241. doi:10.1038/nature21059. Epub 2016 Dec 22

Cass SD, Haas KA, Stoll B, Alkhnbashi O, Sharma K, Urlaub H, Backofen R, Marchfelder A, Bolt EL (2015) The role of Cas8 in type I CRISPR interference. Biosci Rep 5:5

Charlebois RL, Lam WL, Cline SW, Doolittle WF (1987) Characterization of pHV2 from Halobacterium volcanii and its use in demonstrating transformation of an archaebacterium. Proc Natl Acad Sci U S A 84:8530–8534

Charpentier E, Richter H, van der Oost J, White MF (2015) Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. FEMS Microbiol Rev 39:428–441. Epub 2015 May 19

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823 doi:10.1126/science1231143

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logogenerator. Genome Res 14:1188–90

Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun 3:945. doi:10.1038/ncomms1937

Delmas S, Shunburne L, Ngo HP, Allers T (2009) Mre11-Rad50 promotes rapid repair of DNA damage in the polyploid archaeon Haloferax volcanii by restraining homologous recombination. PLoS Genet 5:e1000552. Epub 2009 Jul 10

Deng L, Kenchappa CS, Peng X, She Q, Garrett RA (2012) Modulation of CRISPR locus transcription by the repeat-binding protein Cbp1 in Sulfolobus. Nucleic Acids Res 40:2470–2480. Epub 2011 Dec 1

Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S (2008) Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J Bacteriol 190:1390–1400. Epub 2007 Dec 13

Erdmann S, Garrett RA (2012) Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. Mol Microbiol 85:1044–1056. doi:10.1111/j1365-2958201208171x. Epub 2012 Jul 27

Fineran PC, Dy RL (2014) Gene regulation by engineered CRISPR-Cas systems. Curr Opin Microbiol 18:83–89. doi: 10.1016/j.mib.2014.1002.1007. Epub 2014 Mar 15

Fineran PC, Gerritzen MJ, Suarez-Diez M, Kunne T, Boekhorst J, van Hijum SA, Staals RH, Brouns SJ (2014) Degenerate target sites mediate rapid primed CRISPR adaptation. Proc Natl Acad Sci U S A 111:E1629–E1638. doi:10.1073/pnas.1400071111. Epub 2014 Apr 17

Fischer S, Maier LK, Stoll B, Brendel J, Fischer E, Pfeiffer F, Dyall-Smith M, Marchfelder A (2012) An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. J Biol Chem 287:33351–33365

Garcia-Heredia I, Martin-Cuadrado AB, Mojica FJ, Santos F, Mira A, Anton J, Rodriguez-Valera F (2012) Reconstructing viral genomes from the environment using fosmid clones: the case of haloviruses. PLoS One 7:e33802. doi:10.31371/journalpone0033802. Epub 2012 Mar 30

Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan AH, Moineau S (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468:67–71

Garrett RA, Vestergaard G, Shah SA (2011) Archaeal CRISPR-based immune systems: exchangeable functional modules. Trends Microbiol 19:549–556. Epub 2011 Sep 22

Gleditzsch D, Muller-Esparza H, Pausch P, Sharma K, Dwarakanath S, Urlaub H, Bange G, Randau L (2016) Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. Nucleic Acids Res 44:5872–5882. doi:10.1093/nar/gkw5469. Epub 2016 May 23

Goren MG, Doron S, Globus R, Amitai G, Sorek R, Qimron U (2016) Repeat size determination by two molecular rulers in the Type I-E CRISPR array. Cell Rep 16:2811–2818. doi:10.1016/j.celrep.2016.2808.2043

Haft DH, Selengut J, Mongodin EF, Nelson KE (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol 1:e60. Epub 2005 Nov 11

Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell 139:945–956. doi:10.1016/j.cell.2009.1007.1040

Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, Resch AM, Glover CV, III, Graveley BR, Terns RM, Terns MP (2012) Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. Mol Cell 45:292–302. Epub 2012 Jan 15

Hille F, Charpentier E (2016) CRISPR-Cas: biology, mechanisms and relevance. Philos Trans R Soc Lond B Biol Sci 371(1707):20150496. doi:10.20151098/rstb.20152015.20150496

Hochstrasser ML, Taylor DW, Bhat P, Guegler CK, Sternberg SH, Nogales E, Doudna JA (2014) CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proc Natl Acad Sci U S A 111:6618–6623. doi:10.1073/pnas.1405079111. Epub 2014 Apr 18

Ivancic-Bace I, Cass SD, Wearne SJ, Bolt EL (2015) Different genome stability proteins underpin primed and naive adaptation in E. coli CRISPR-Cas immunity. Nucleic Acids Res 43:10821–10830. doi:10.11093/nar/gkv11213. Epub 2015 Nov 17

Jackson RN, Golden SM, van Erp PB, Carter J, Westra ER, Brouns SJ, van der Oost J, Terwilliger TC, Read RJ, Wiedenheft B (2014) Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. Science 345:1473–1479

Jansen R, Embden JD, Gaastra W, Schouls LM (2002a) Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol 43:1565–1575

Jansen R, van Embden JD, Gaastra W, Schouls LM (2002b) Identification of a novel family of sequence repeats among prokaryotes. OMICS 6:23–33

Jiang W, Maniv I, Arain F, Wang Y, Levin BR, Marraffini LA (2013) Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. PLoS Genet 9: e1003844. doi:10.1001371/journalpgen1003844. Epub 2013 Sep 26

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821. doi:10.1126/science.1225829. Epub 2012 Jun 28

Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul Ü, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol 18:529–536

Kunin V, Sorek R, Hugenholtz P (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol 8:R61

Künne T, Swarts DC, Brouns SJ (2014) Planting the seed: target recognition of short guide RNAs. Trends Microbiol 22:74–83. doi:10.1016/j.tim.2013.1012.1003. Epub 2014 Jan 14

Künne T, Kieper SN, Bannenberg JW, Vogel AI, Miellet WR, Klein M, Depken M, Suarez-Diez M, Brouns SJ (2016) Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. Mol Cell 63:852–864. doi:10.1016/j.molcel.2016.1007.1011. Epub 2016 Aug 18

Labrie SJ, Samson JE, Moineau S (2010) Bacteriophage resistance mechanisms. Nat Rev Microbiol 8:317–327. Epub 2010 Mar 29

Lange SJ, Alkhnbashi OS, Rose D, Will S, Backofen R (2013) CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. Nucleic Acids Res 41:8034–8044

Levy A, Goren MG, Yosef I, Auster O, Manor M, Amitai G, Edgar R, Qimron U, Sorek R (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature 520:505–510. doi: 10.1038/nature14302. Epub 2015 Apr 13

Li H (2015) Structural Principles of CRISPR RNA Processing. Structure 23:13–20. doi:10.1016/j.str.2014.1010.1006. Epub 2014 Nov 26

Li M, Liu H, Han J, Liu J, Wang R, Zhao D, Zhou J, Xiang H (2013) Characterization of CRISPR RNA biogenesis and Cas6 cleavage-mediated inhibition of a provirus in the haloarchaeon Haloferax mediterranei. J Bacteriol 195:867–875. doi:10.1128/JB.01688-01612. Epub 2012 Dec 14

Li M, Wang R, Xiang H (2014a) Haloarcula hispanica CRISPR authenticates PAM of a target sequence to prime discriminative adaptation. Nucleic Acids Res 42:7226–7235. doi:10.1093/nar/gku7389. Epub 2014 May 26

Li M, Wang R, Zhao D, Xiang H (2014b) Adaptation of the Haloarcula hispanica CRISPR-Cas system to a purified virus strictly requires a priming process. Nucleic Acids Res 42:2483–2492. doi:10.1093/nar/gkt1154. Epub 2013 Nov 21

Luo ML, Mullis AS, Leenay RT, Beisel CL (2015) Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. Nucleic Acids Res 43:674–681. doi:10.1093/nar/gku1971. Epub 2014 Oct 17

Maier LK, Fischer S, Stoll B, Brendel J, Pfeiffer F, Dyall-Smith M, Marchfelder A (2012) The immune system of halophilic archaea. Mob Genet Elements 2:228–232

Maier LK, Lange SJ, Stoll B, Haas KA, Fischer S, Fischer E, Duchardt-Ferner E, Wohnert J, Backofen R, Marchfelder A (2013a) Essential requirements for the detection and degradation of invaders by the Haloferax volcanii CRISPR/Cas system I-B. RNA Biol 10:865–874

Maier LK, Stoll B, Brendel J, Fischer S, Pfeiffer F, Dyall-Smith M, Marchfelder A (2013b) The ring of confidence: a haloarchaeal CRISPR/Cas system. Biochem Soc Trans 41:374–378. doi:10.1042/BST20120263

Maier LK, Dyall-Smith M, Marchfelder A (2015a) The adaptive immune system of Haloferax volcanii. Life (Basel) 5:521–537. doi:10.3390/life5010521

Maier LK, Stachler AE, Saunders SJ, Backofen R, Marchfelder A (2015b) An active immune defense with a minimal CRISPR (clustered regularly interspaced short palindromic repeats) RNA and without the Cas6 protein. J Biol Chem 290:4192–4201. doi:10.1074/jbc.M4114.617506. Epub 2014 Dec 15

Makarova KS, Aravind L, Wolf YI, Koonin EV (2011) Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biol Direct 6:38. doi:10.1186/1745-6150-1186-1138

Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJ, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV (2015) An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol 13:722–736. doi:10.1038/nrmicro3569. Epub 2015 Sep 28

Marraffini LA, Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322:1843–1845

Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin EV, van der Oost J (2016) Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science 353:aad5147. doi:10.1126/science.aad5147

Mojica FJ, Rodriguez-Valera F (2016) The discovery of CRISPR in archaea and bacteria. FEBS J 283:3162–3169. doi:10.1111/febs.13766. Epub 2016 Jun 15

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol 60:174–182

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155:733–740

Mulepati S, Heroux A, Bailey S (2014) Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. Science 345:1479–1484

Nickel L, Weidenbach K, Jäger D, Backofen R, Lange SJ, Heidrich N, Schmitz RA (2013) Two CRISPR-Cas systems in Methanosarcina mazei strain Göo1 display common processing features despite belonging to different types I and III. RNA Biol 10:779–791

Niewoehner O, Jinek M, Doudna JA (2014) Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. Nucleic Acids Res 42:1341–1353. doi:10.1093/nar/gkt1922. Epub 2013 Oct 22

Norais C, Hawkins M, Hartman AL, Eisen JA, Myllykallio H, Allers T (2007) Genetic and physical mapping of DNA replication origins in Haloferax volcanii. PLoS Genet 3:e77. Epub 2007 Apr 5

Nunez JK, Harrington LB, Kranzusch PJ, Engelman AN, Doudna JA (2015) Foreign DNA capture during CRISPR-Cas adaptive immunity. Nature 527:535–538. doi:10.1038/nature15760. Epub 2015 Oct 21

Oren A (2006) The order halobacteriales. In: Dworkin M, DeLong EF, Stackebrandt E, Lory S, Thompson F (eds) The prokaryotes, vol 3. Springer, New York, pp 113–164

Plagens A, Tjaden B, Hagemann A, Randau L, Hensel R (2012) Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon Thermoproteus tenax. J Bacteriol 194:2491–2500. doi:10.1128/JB.00206-00212. Epub 2012 Mar 9

Pourcel C, Salvignol G, 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

Pul Ü, Wurm R, Arslan Z, Geissen R, Hofmann N, Wagner R (2010) Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. Mol Microbiol 75:1495–1512

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152:1173–1183. doi:10.1016/j.cell.2013.1102.1022

Rath D, Amlinger L, Hoekzema M, Devulapally PR, Lundgren M (2015) Efficient programmable gene silencing by Cascade. Nucleic Acids Res 43:237–246. doi:10.1093/nar/gku1257. Epub 2014 Nov 30

Richter H, Zoephel J, Schermuly J, Maticzka D, Backofen R, Randau L (2012) Characterization of CRISPR RNA processing in Clostridium thermocellum and Methanococcus maripaludis. Nucleic Acids Res 40:9887–9896. doi:10.1093/nar/gks9737. Epub 2012 Aug 8

Richter C, Dy RL, McKenzie RE, Watson BN, Taylor C, Chang JT, McNeil MB, Staals RH, Fineran PC (2014) Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. Nucleic Acids Res 42:8516–8526. doi:10.1093/nar/gku8527. Epub 2014 Jul 12

Sampson TR, Weiss DS (2014) Exploiting CRISPR/Cas systems for biotechnology. Bioessays 36:34–38. doi:10.1002/bies.201300135

Sashital DG, Wiedenheft B, Doudna JA (2012) Mechanism of foreign DNA selection in a bacterial adaptive immune system. Mol Cell 46:606–615. doi:101016/jmolcel201210031020. Epub 2012 Apr 19

Scholz I, Lange SJ, Hein S, Hess WR, Backofen R (2013) CRISPR-Cas systems in the cyanobacterium Synechocystis sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein. PLoS One 8:e56470. doi:10.51371/journalpone0056470. Epub 2013 Feb 18

Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJ, 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. Epub 2011 Jun 6

Semenova E, Savitskaya E, Musharova O, Strotskaya A, Vorontsova D, Datsenko KA, Logacheva MD, Severinov K (2016) Highly efficient primed spacer acquisition from targets destroyed by the Escherichia coli type I-E CRISPR-Cas interfering complex. Proc Natl Acad Sci U S A 113:7626–7631. doi:10.1073/pnas.1602639113. Epub 2016 Jun 20

Shah SA, Garrett RA (2011) CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. Res Microbiol 162:27–38. Epub 2010 Sep 21

Shah SA, Erdmann S, Mojica FJ, Garrett RA (2013) Protospacer recognition motifs: mixed identities and functional diversity. RNA Biol 10:5

Shao Y, Li H (2013) Recognition and cleavage of a nonstructured CRISPR RNA by its processing endoribonuclease Cas6. Structure 21:385–393

Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, Minakhin L, Joung J, Konermann S, Severinov K, Zhang F, Koonin EV (2015) Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. Mol Cell 60:385–397. doi:101016/jmolcel201510101008. Epub 2015 Oct 22

Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, Abudayyeh OO, Gootenberg JS, Makarova KS, Wolf YI, Severinov K, Zhang F, Koonin EV (2017) Diversity and evolution of class 2 CRISPR-Cas systems. Nat Rev Microbiol 15:169–182. doi:10.1038/nrmicro.2016.184. Epub 2017 Jan 23

Staals RH, Brouns SJ (2013) Distribution and mechanism of the type I CRISPR-Cas systems. In: Barrangou R, van der Oost J (eds) CRISPR-Cas systems. Springer, Berlin, pp 145–169

Staals RH, Jackson SA, Biswas A, Brouns SJ, Brown CM, Fineran PC (2016) Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. Nat Commun 7:12853. doi:10.1038/ncomms12853

Stachler AE, Marchfelder A (2016) Gene repression in Haloarchaea using the CRISPR (clustered regularly interspaced short palindromic repeats) – Cas I-B system. J Biol Chem 291:15226–15242

Sternberg SH, Doudna JA (2015) Expanding the biologist's toolkit with CRISPR-Cas9. Mol Cell 58:568–574. doi:10.1016/j.molcel.2015.1002.1032

Sternberg SH, Richter H, Charpentier E, Qimron U (2016) Adaptation in CRISPR-Cas systems. Mol Cell 61:797–808. doi:10.1016/jmolcel201610011030. Epub 2016 Mar 13

Stoll B, Maier LK, Lange SJ, Brendel J, Fischer S, Backofen R, Marchfelder A (2013) Requirements for a successful defence reaction by the CRISPR-Cas subtype I-B system. Biochem Soc Trans 41:1444–1448. doi:10.1042/BST20130098

Suttle CA (2007) Marine viruses–major players in the global ecosystem. Nat Rev Microbiol 5:801–812

Swarts DC, Mosterd C, van Passel MW, Brouns SJ (2012) CRISPR interference directs strand specific spacer acquisition. PLoS One 7:e35888. doi:10.31371/journalpone0035888. Epub 2012 Apr 27

Vestergaard G, Garrett RA, Shah SA (2014) CRISPR adaptive immune systems of Archaea. RNA Biol 11:156–167. doi:10.4161/rna.27990. Epub 2014 Feb 7

Vorontsova D, Datsenko KA, Medvedeva S, Bondy-Denomy J, Savitskaya EE, Pougach K, Logacheva M, Wiedenheft B, Davidson AR, Severinov K, Semenova E (2015) Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery. Nucleic Acids Res 43:10848–10860. doi:10.11093/nar/gkv11261. Epub 2015 Nov 19

Wang R, Zheng H, Preamplume G, Shao Y, Li H (2012) The impact of CRISPR repeat sequence on structures of a Cas6 protein-RNA complex. Protein Sci 21:405–417. doi:10.1002/pro.2028. Epub 2012 Feb 9

Wang J, Li J, Zhao H, Sheng G, Wang M, Yin M, Wang Y (2015) Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. Cell 163:840–853. doi:10.1016/jcell201510101008. Epub 2015 Oct 17.

Wang R, Li M, Gong L, Hu S, Xiang H (2016) DNA motifs determining the accuracy of repeat duplication during CRISPR adaptation in Haloarcula hispanica. Nucleic Acids Res 44:4266–4277. doi:10.1093/nar/gkw4260. Epub 2016 Apr 16

Westra ER, Semenova E, Datsenko KA, Jackson RN, Wiedenheft B, Severinov K, Brouns SJ (2013) Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. PLoS Genet 9:e1003742. doi:10.1001371/journalpgen1003742. Epub 2013 Sep 5

Wiedenheft B, van Duijn E, Bultema JB, Waghmare SP, Zhou K, Barendregt A, Westphal W, Heck AJ, Boekema EJ, Dickman MJ, Doudna JA (2011) RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. Proc Natl Acad Sci U S A 108:10092–10097. Epub 2011 May 2

Woods WG, Dyall-Smith ML (1997) Construction and analysis of a recombination-deficient (radA) mutant of Haloferax volcanii. Mol Microbiol 23:791–797

Yosef I, Goren MG, Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res 40:5569–5576. doi:10.1093/nar/gks5216. Epub 2012 Mar 8

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, Koonin EV, Zhang F (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163:759–771. doi:10.1016/j.cell.2015.1009.1038. Epub 2015 Sep 25

Zhang J, Rouillon C, Kerou M, Reeks J, Brugger K, Graham S, Reimann J, Cannone G, Liu H, Albers SV, Naismith JH, Spagnolo L, White MF (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. Mol Cell 45:303–313. doi:10.1016/j.molcel.2011.1012.1013. Epub 2012 Jan 15

Zhao H, Sheng G, Wang J, Wang M, Bunkoczi G, Gong W, Wei Z, Wang Y (2014) Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. Nature 515(7525):147–150