RNA Biology 10:5, 891–899; May 2013; © 2013 Landes Bioscience

Protospacer recognition motifs

Mixed identities and functional diversity

Shiraz A. Shah,[1] Susanne Erdmann,[1] Francisco J.M. Mojica[2] and Roger A. Garrett[1,*]

[1]Archaea Centre; Department of Biology; University of Copenhagen; Copenhagen, Denmark; [2]Departamento de Fisiología, Genetica y Microbiología; Facultad de Ciencias; Universidad de Alicante; Alicante, Spain

Protospacer adjacent motifs (PAMs) were originally characterized for CRISPR-Cas systems that were classified on the basis of their CRISPR repeat sequences. A few short 2–5 bp sequences were identified adjacent to one end of the protospacers. Experimental and bioinformatical results linked the motif to the excision of protospacers and their insertion into CRISPR loci. Subsequently, evidence accumulated from different virus- and plasmid-targeting assays, suggesting that these motifs were also recognized during DNA interference, at least for the recently classified type I and type II CRISPR-based systems. The two processes, spacer acquisition and protospacer interference, employ different molecular mechanisms, and there is increasing evidence to suggest that the sequence motifs that are recognized, while overlapping, are unlikely to be identical. In this article, we consider the properties of PAM sequences and summarize the evidence for their dual functional roles. It is proposed to use the terms protospacer associated motif (PAM) for the conserved DNA sequence and to employ spacer acqusition motif (SAM) and target interference motif (TIM), respectively, for acquisition and interference recognition sites.

Keywords: adaptive immunity, CRISPR, protospacer, PAM, SAM, TIM

Submitted: 12/19/12

Revised: 01/21/13

Accepted: 01/24/13

http://dx.doi.org/10.4161/rna.23764

*Correspondence to: Roger A. Garrett; Email: garrett@bio.ku.dk

Introduction

Clustered, regularly interspaced, short palindromic repeats (CRISPR) provide a basis for the disparate adaptive immune systems that occur in most archaea and many bacteria.[1-5] First insights into the function of these CRISPR arrays arose from the discovery that sequences of some CRISPR spacer regions closely matched sequences occurring in viruses or plasmids. This led, in turn, to the proposal that they participate in defense against invading genetic elements.[6-8] This hypothesis was subsequently supported by experiments showing that newly acquired spacers deriving from a group of bacteriophages produced viral immunity in strains of *Streptococcus thermophilus*.[9-11] These seminal developments constituted a major breakthrough in microbiology.

Spacers derive from fragments of invading genetic elements termed protospacers, and they are incorporated into CRISPR loci generally, but not invariably, at repeats adjacent to CRISPR leaders.[9-14] CRISPR loci are transcribed from the leader and processed within repeats to yield small crRNAs, carrying most or all of the spacer sequence. crRNAs act as guide RNAs for different interference modules that target and cleave DNA or RNA after annealing to the complementary protospacer sequence within nucleic acid of the invading element.[15-20]

A DNA sequence element that is functionally critical for CRISPR-based immune systems is located adjacent to each protospacer. It consists of a short signature sequence of 2–5 bp that varies according to the CRISPR-based system and organism. This motif was first detected in sequence alignments of putative protospacers of bacteriophages that match CRISPR spacers of *Streptococcus* strains,[8] and subsequently other diverse motifs were defined for a variety of organisms and different types of CRISPR systems.[9,11,18,21-23] The

demonstration of the apparent universality of the short sequence motifs adjoining protospacers by Mojica et al.[21] led to their assigning the acronym PAM for protospacer adjacent motif.

CRISPR-based systems have recently been reclassified into three main types, I, II and III, where the former two types target DNA elements while the type III systems target either DNA or RNA.[3] PAM sequences of protospacers incorporated into CRISPR loci associated with type I systems are located at the protospacer end that becomes leader proximal,[18,21-23] whereas those acquired by CRISPR loci associated with the bacteria-specific type II CRISPR-Cas systems occur at the leader distal end.[8-11] The first evidence implicating the PAM sequence in the interference mechanism was provided for the type II-A system of *S. thermophilus*,[9,11] and this result was reinforced by interference experiments on other CRISPR-based systems.[23-29]

In this article, we focus attention on the PAM sequence and reexamine its potential functional versatility. Currently, PAMs have been predicted for CRISPR-based systems of a few organisms for which several compatible genetic elements have been sequenced and that carry a significant number of identifiable protospacers.[8,11,18,21,22] Therefore, we focus mainly on the following CRISPR types and organisms from which many of the seminal bioinformatical and experimental results derive: type I-A of members of the archaeal order Sulfolobales, type I-E and I-F of *Escherichia coli* and type II-A of *S. thermophilus*. For the Sulfolobales, genome sequences are available for several organisms which carry large and complex CRISPR loci as well as being hosts for many diverse viruses and plasmids.[14,18,24,30] *E. coli* carries the relatively simple and streamlined type I-E and I-F systems that are particularly amenable to genetic, biochemical and structural analyses,[12,13,19,23,28,31-33] while *S. thermophilus* contains a bacteria-specific type II-A system that yielded some of the first insights into the PAM dependence of spacer acquisition and protospacer interference.[9-11] Surveys of genetic elements of other laboratory strains of archaea and bacteria have yielded relatively few reliable PAM sequences.[21,26,34]

In this article, we examine the experimental evidence for the functional roles of the PAM sequence in spacer acquisition and interference and propose defining two distinct functional motifs, a spacer acquisition motif (SAM) for acquisition and a target interference motif (TIM) for interference.

Characterization of Protospacer-Associated Sequences

The first evidence for a conserved sequence motif adjacent to predicted protospacers on phages and plasmids was found for a type II-A CRISPR system of *S. thermophilus.* Protospacer alignments revealed a degenerate sequence 5'-NNpu-py-A-A-a-3' downstream from several putative protospacers.[8] The authors implied that the similarity of the sequence to the conserved terminal repeat sequence ACAAC, except for the terminal nucleotide, might be mechanistically significant. However, subsequent studies on different *S. thermophilus* strains revealed two motifs, NNAGAAa and NGGNG, associated with bacteriophage protospacers that were actively acquired as spacers in two coexisting type II systems. However, the NGGNG motif showed no similarity to the terminal sequence AAAAC of the CRISPR repeat.[9]

These developments coincided with the first attempts by Kunin et al.[35] to classify CRISPR-based systems on the basis of the sequences and inverted repeat contents of CRISPR repeats. They defined 12 main families, some of which showed a distinct phylogenetic bias, in particular to archaea or bacteria.[35] Mojica et al.[21] examined members of these repeat-based CRISPR families looking for characteristic consensus sequence motifs associated with predicted protospacers. Although the approach was limited by the difficulty in predicting multiple protospacers in genetic elements for most CRISPR-containing archaea and bacteria, significant numbers of protospacers were found to match CRISPR loci of a few organisms belonging to different repeat families. These consensus protospacer adjacent motifs were then assigned the acronym PAM.[21] The main PAM assignments from this study are summarized in **Table 1** together with more recent results. Here, we present the sequence at the 5'-end of the crRNA sense strand of the protospacer (5'-PAM-protospacer) for the type I systems and the opposite orientation 5'-protospacer-PAM for the type II systems. However, there are still relatively few CRISPR-carrying organisms for which reliable PAM sequences have been identified.

Putative Role in Spacer Acquisition

The acquisition process involves recognition and excision of protospacers and their insertion into CRISPR loci, and it appears to be the most conserved stage of the adaptive immune response. Generally, three proteins, Cas1, Cas2 and Cas4, have been implicated,[4] although the type I-E and I-F systems of *E. coli* lack Cas4, as do type II-A and some type III-A systems.[3] The largest and most conserved protein, Cas1, carries DNA endonuclease activity,[36,37] and in an *E. coli* type I-E system, its mutation can inhibit spacer acquisition.[12] Cas2 protein from *Bacillus halodurans* also exhibits dsDNA endonuclease activity,[38] while another Cas2 protein from *Sulfolobus solfataricus* and other archaea, showed low specificity ssRNA endonuclease activity, currently of uncertain biological significance.[39,40] Cas4 of *S. solfataricus* carries 5'- to 3'-DNA exonuclease activity that may generate recombigenic 3'-overlaps for CRISPR spacer insertion.[41]

A potential link between acquisition and the PAM sequence was provided earlier for the Sulfolobales by Shah et al.[18,22] Distance trees were prepared for many different CRISPR loci of several members of the Sulfolobales based on sequences of CRISPR repeats, leaders and Cas1 proteins. Each tree showed three similar major branches containing components (repeat, leader or Cas1 protein) associated with the same CRISPR loci. Moreover, prediction of sequence motifs (PAMs) adjoining putative protospacers exhibiting significant sequence matches to spacers within the different CRISPR loci revealed strong biases to CCN, TCN and GTN, respectively, for the three main branches.[18,22] This suggested that in addition to Cas1, the protospacer motif, repeat and leader were involved in acquisition. These results are updated in **Figure 1A** for the repeat and

Table 1. Summary of experimental data relating to the dependence of DNA interference on PAM

Organism	Subtype	PAM	Interference (+)	Interference (-)	Reference
S. solfataricus	I-A1	CCN	CCA TCA	GAC TTA	18, 24
S. solfataricus	I-A2	TCN*	TCG		30
H. volcanii	I-B	n.d.	TTC ACT TAA TAT TAG CAC	remaining 58 trinucleotides	26
H. walsbyi	I-B	TTC*	n.d.		56
E. coli	I-E	AWG	ATG AAG GAG AGG	ACG CAG TAG GTG TGG TTG CCG CTG AAA AAC AAT AGC ATA ATC ATT GAA CAA CCT CCC GGC TCC TCT	13, 23, 32, 33
E. coli	I-F	CC	GCC CCC GCT CTT CAA	GTC AAA GTT GGG ACA GCA AAT AAC	28
P. aeruginosa	I-F	CC		AG	58
S. thermophilus	II-A	NNAGAA	AGAA	AGAG AAAA ATAA	11
S. agalactiae	II-A	NGG	NGG NGA	NAC NCA NAG NGT	27
S. thermophilus	II-A	NGGNG	GGNG	CGTG GCTG GGTC	25

CRISPR subtypes are given for each organism. *, indicates that the PAM was determined from sequence alignments. Interference dependence on permutations of the PAM are summarized in the interference columns where (+) indicates successful interference and (-) denotes no or little interference. All triplet sequences are drawn 5' to 3'. Literature references describing the original results are provided. n.d., not determined.

Cas1 protein sequences, including many new sequences. The tree reveals that the *cas1* gene of a given CRISPR subtype always occurs together with genomic CRISPR arrays of the same subtype. A single exception is *Metallosphaera cuprina*, which probably results from the occurrence of an IS element-mediated transposition between the *cas1* gene and the adjacent CRISPR array. Furthermore, logoplots of the predicted PAMs reinforce that there is a close correlation between the sequence identity of PAM and the CRISPR subtype (**Fig. 1B**). These results reinforce and extend the earlier sequence analyses[18,22] demonstrating the strong interdependence of the type of CRISPR array, the Cas1 protein and PAM and they correlate with the contemporary evidence for coevolution of the PAM sequence and CRISPR repeat families (**Table 1**).[21,34]

Further support for the involvement of the leader in acquisition came from the observation that no spacer uptake was observed in the leaderless CRISPR locus F that is highly conserved in sequence between different *S. solfataricus* strains, whereas all other CRISPR loci carrying leaders acquired new spacers.[14,17,18] More direct evidence for a leader role in acquisition was provided recently by Yosef et al.[12] who demonstrated that an unknown sequence located within the first 60 bp of the leader, adjacent to the first CRISPR repeat, was essential for spacer acquisition in an *E. coli* type I-E system and the size of this important region was further reduced to 43 bp in an independent study on this type I-E system.[42]

At present, little is known about the detailed mechanisms of spacer acquisition. The PAM sequence is likely to generate a recognition site for type I protospacer excision from genetic elements, or fragments thereof, with cleavage occurring adjacent to the PAM sequence (see below). At the other end of the protospacer there is no detectable sequence specificity for cleavage.[14] Moreover, multiple cutting sites can occur over up to six base pairs for a given protospacer region in different copies of the same genetic element.[14] This led Erdmann and Garrett[14] to propose that a ruler cleavage mechanism occurs for protospacer excision measured from the PAM sequence. Such a mechanism is also consistent with the observation that many protospacers contain internal PAM sequences that can also be recognized, independently, during protospacer excision from other copies of the same genetic element.[14] Díez-Villaseñor et al.[42] have also proposed a second ruler mechanism operating at the spacer insertion stage whereby an initial cleavage occurs at the leader-repeat boundary of a CRISPR locus with a secondary cut occurring at the leader distal end of the first repeat.[42] This shared ruler strategy during the two main acquisition steps could ensure maintenance of the regular periodicity within CRISPR loci.[42]

Evidence has also been presented for the occasional uptake of spacers in a reverse direction, for a type I-A system of *S. solfataricus*,[14] a type II-A system of *Streptococcus agalactiae*[27] and a type I-E system of *E. coli*,[42] and this places some constraints on possible mechanisms of spacer insertion.[43] Moreover, it may be significant for understanding details of the insertion mechanisms employed in *S. solfataricus* and *Streptococcus agalactiae* type I-A and type II-A systems, respectively, that PAM sequences located at opposite ends of the protospacer generate the same sequences when inverted 5'-CCN vs. NGG-3'.

A different picture is emerging from recent studies on a genetically manipulated *E. coli* type I-E system that carries a single gene cassette encoding acquisition and interference Cas proteins.[44-46] Yosef et al.[12] first induced spacer acquisition by overexpressing proteins Cas1 and Cas2, and found relatively low conservation at positions -3 and -2 in the AWG PAM, whereas position -1G was highly conserved. They also provided evidence for the new repeat being copied from repeat 1 during spacer acquisition. Swarts et al.[13] demonstrated examples of compensatory base changes between the -1 PAM position and the terminal downstream nucleotide of the repeat at which spacer insertion occurred. They inferred that the PAM -1 nucleotide is taken up in the newly synthesized repeat, a conclusion that was supported by Datsenko et al.[32] Goren et al.[47] took this a stage further proposing that the -1 position of the PAM should be considered the terminal nucleotide of the protospacer sequence and

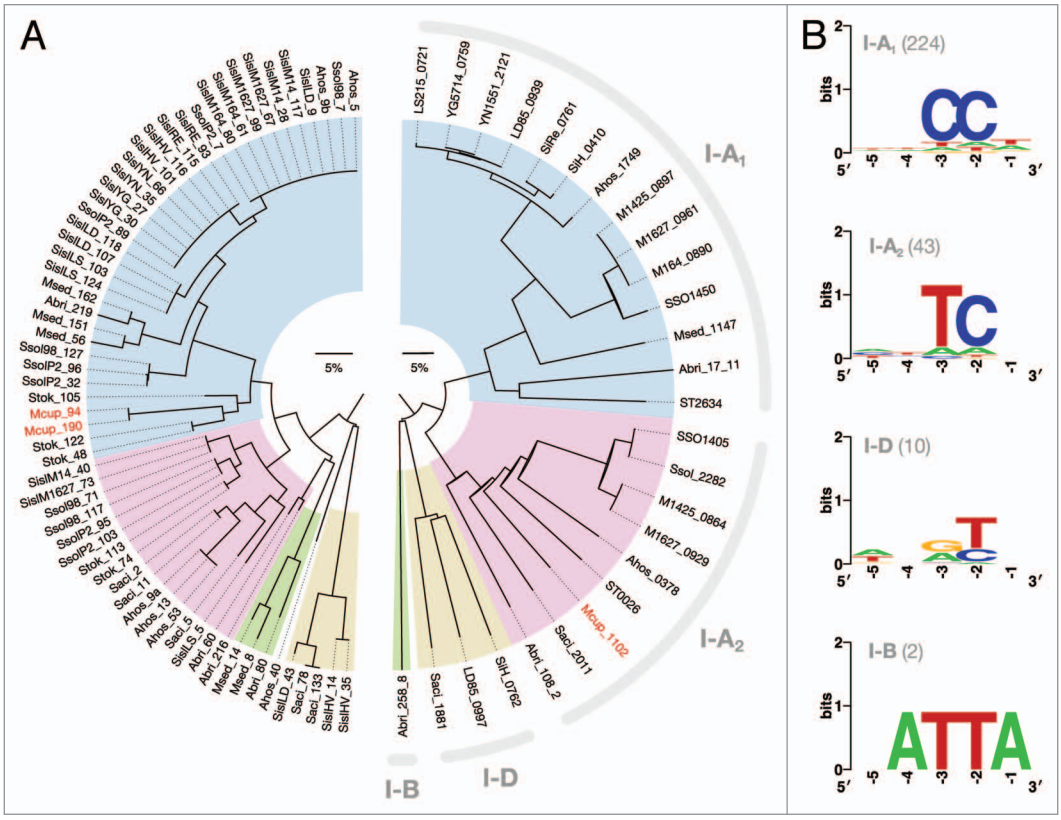

Figure 1. Coevolution of sequences of CRISPR repeats, Cas1 proteins and PAMs, for members of the Sulfolobales. The type I CRISPR systems fall into three main subtypes, I-A, I-D and I-B. The I-A systems are the most common and can be classified into distinct subfamilies I-A_1 and I-A_2. (**A**) A neighbor joining tree of CRISPR repeat sequences (left) is juxtaposed with that of the translated sequences of *cas1* genes (right). CRISPR loci are identified by the short name of the organism and the number of CRISPR repeats. *cas1* genes are colored according to the CRISPR subtypes (I-A_1, blue; I-A_2, pink; I-D, yellow and I-B, green). (**B**) Protospacer matches from spacers of the four distinct subtypes of CRISPR arrays yield dominant consensus PAMs, CCN for I-A_1, TCN for I-A_2, GTN for I-D and with ATTA for two protospacers predicted for subtype I-B. Motifs were derived from spacer-protospacer matches on viral or plasmid genomes of the Sulfolobales exhibiting five or less mismatches. The total number of predicted protospacers is given in brackets.

not a part of the repeat. Mojica et al.[21] had earlier identified CRISPR systems of other organisms for which the -1 PAM position is conserved, including some type I-C, I-B and I-F systems. For some of them, the -1 PAM position matches the first nucleotide of the repeat and could also, potentially, be assigned to the protospacer.[21]

In summary, whereas the results for the type I-E system of *E. coli* provide some insights into how specific spacer insertion into CRISPR repeats can occur by exploiting the -1 position of the PAM sequence, this mechanism cannot be generally applicable because for many type I and type II systems, the equivalent PAM position is not conserved.[14,21]

PAM-Protospacer Selection on Genetic Elements

It was estimated from statistical analyses of distributions of predicted protospacers, in diverse viral and plasmid genomes of the Sulfolobales, that they were located randomly on both circular and linear genomes.[22,48] They exhibited no significant bias with respect to either direction or to their location within protein coding or non-coding regions. This study also provided support for the PAM sequence being independent of the type of genetic element from which the protospacer originated because a high incidence of putative protospacers from both linear viral genomes and circular (inferred to be positively supercoiled[49]) plasmid and viral genomes occurred within the same CRISPR loci.[18] In an extensive study of bacteria and archaea, Mojica et al.[21] reached similar conclusions for a wide range of genetic elements.

Recently, hundreds of spacers acquired from a conjugative plasmid co-infecting *S. solfataricus* with a tailed-fusiform virus were sequenced and analyzed for their PAM sequences.[14] Originally, these results were presented for a few smaller, unlinked, contigs but here, 399 unique protospacers are reanalysed for a large 22 kb contig of the plasmid from Monument Geysir Basin-Yellowstone National Park (herein named pMGB1), and the data are presented in **Table 2**. The results quantify the high level of conservation of the acquisition CCN PAM sequences (95%). Moreover, they reaffirm the conclusion from the earlier bioinformatical analyses that protospacers occur randomly throughout the targeted genetic element on both strands with no significant bias to predicted protein coding regions.[14,22,48] The random distribution of acquired protospacers is also consistent with the results obtained for *Streptococcus* type II systems[9,11,50] and the genetically modified *E. coli* type I-E system.[12,13]

There are, nevertheless, significant differences between results obtained for the type I-A system of *Sulfolobus* and the type I-E system of *E. coli*. First, the -1 PAM position is only conserved for the latter system while the -2 and -3 PAM positions for the type I-A system of *Sulfolobus* are much more highly conserved than in the type I-E system.[12-14] Second, when multiple protospacers are incorporated into a CRISPR locus of a single clone they derive from unidirectional protospacers on a genetic element for the *E. coli* type I-E system but this was not observed for the Sulfolobales.[14] This is exemplified by an in silico analysis of the type I-A system of *Metallosphaera sedula*, where the orientations of 30 putative protospacers matching (with ≤ 5 nt mismatches) the genome of the *Acidianus* two-tailed virus, ATV, are arranged bidirectionally for sections of three CRISPR loci (**Fig. 2**).[13,14,32]

Datsenko et al.[32] provided a rationale for unidirectional uptake of protospacers within a specific CRISPR locus of the type I-E system. They hypothesized that low level annealing of crRNAs from older spacers with newly invading genetic elements can stimulate ("prime") acquisition of new spacers from that element. This implicit coupling of interference and acquisition is indirectly supported by the co-expression of the acquisition proteins Cas1 and Cas2 and interference-related Cas proteins in that type I-E system. Moreover, the missing Cas4 DNA exonuclease,[40] which is implicated in acquisition for other CRISPR types,[46] may be complemented by the interference Cas3 endonuclease.[32] Consistent with this proposal, in the type I-F system of *Pectobacterium atrosepticum*, it has been shown that Cas1 interacts with a Cas2-Cas3 hybrid protein.[51] A major advantage of reverse coupling of interference and spacer acquisition[32] could be that it provides a means of preferentially selecting genetic elements for interference and thereby facilitates avoidance of chromosomal interference, although priming by a low level of crRNA base pairing could still lead to fortuitous targeting of chromosomal sites.

If this hypothesis were more generally applicable to other CRISPR acquisition systems, it could resolve an earlier puzzle as to why individual CRISPR loci often carry multiple spacer matches against single genetic elements.[17,18] Although it has been demonstrated experimentally that acquisition of more than one spacer from an invading genetic element can provide increased immunity against that element for both the type II-A system of *S. thermophilus*[9,11] and the type I-E system of *E. coli*,[12] CRISPR loci often carry many spacers matching a given element.[17,18,52,53] This phenomenon is exemplified by the 30 spacers predicted to match the lytic virus ATV[54] within three CRISPR loci of the crenarchaeon *M. sedula* (**Fig. 2**). Another explanation for multiple CRISPR spacers matching a single genetic element, especially for the crenarchaea where many viruses coexist in stable relationships with their hosts, is the possibility that CRISPR systems adopt a regulatory role by exhibiting limited levels of interference.[17,18] Almendros et al.[28] have also recently emphasized the cellular disadvantages of CRISPR-based systems being too efficient and rejecting potentially beneficial foreign DNA.

Protospacer Recognition during Interference

While there is strong support for the involvement of PAM sequences in the spacer acquisition step, their importance for interference is less clear. Barrangou et al.[9,11,55] provided the first evidence implicating PAM sequences in this stage of the immune response. They demonstrated for the type II-A system of *S. thermophilus* that mutations in the AGAA motif, located downstream from the protospacer, allowed phages to avoid CRISPR defense. Moreover, experiments involving targeting of plasmid protospacers in *Sulfolobus* provided support for a PAM sequence role in interference.[24] In the presence of the CCN PAM sequence, interference was effective with the few surviving transformants primarily carrying deletions in CRISPR loci that included the matching spacer. When the PAM sequence was replaced with GGN, GAN or TTN, there was no detectable interference. However, in the presence of TCN, (and to a lesser degree CTN) there was a significant reduction in transformation efficiency consistent with an intermediate level of targeting. This contrasted with the stringent acquisition PAM sequence-dependence (**Table 2**) and suggested an altered mode of PAM recognition occurring during interference, where possibly a C, at position -2 or -3 was sufficient (**Table 1**).[12-14]

Table 2. A summary of protospacer acquisition results from 399 sequenced non identical protospacers on a 22 kb contig of the *Sulfolobus* conjugative plasmid pMGB1 by subfamily I-A_i CRISPR loci C, D and E in *S. solfataricus* P2

Protospacer properties	Protospacers (%)
forward	52
reverse	48
CCN PAM	95
"inverted" CCN PAM	0.5
no PAM	4.5

The data are derived from an experimental study by Erdmann and Garrett.[14] The designations "forward" and "reverse" are arbitrary. "No PAM" includes a variety of different dinucleotide sequences, including CTN and TCN.

The first systematic analysis of the interference motif was performed on the haloarchaeon *Haloferax volcanii*, employing a similar plasmid-targeting approach to *Sulfolobus*.[26] Although no specific PAM sequence has been identified for the type I-B CRISPR-Cas system of this organism, it was demonstrated that in total, six protospacer-adjacent triplets TTC, ACT, TAA, TAT, TAG and CAC, out of the 64 possible triplets tested, rendered protospacers active for targeting.[26] None of the triplet positions is completely conserved, and only positions -2 and -3 show limited conservation, suggesting that the PAM sequence might be TAN. Moreover, TTC PAM has been predicted for the type I-B system of another haloarchaeon, *Haloquadratum walsbyi*.[56] In summary, these results on the archaeal type I-A and I-B systems are consistent with different PAM recognition at the DNA interference stage of type I systems, possibly limited to one nucleotide, with some sequence permutations being permitted.

A similar conclusion was recently reached by Lopez-Sanchez et al.[27] for a type II-A system of *S. agalactiae*, where the efficiency of transformation of plasmid constructs carrying protospacers was also studied. Whereas no transformants were formed when the downstream PAM NGG was present, no interference occurred when the dinucleotide was converted to AC, CA, AG or GT. However, the dinucleotide GA

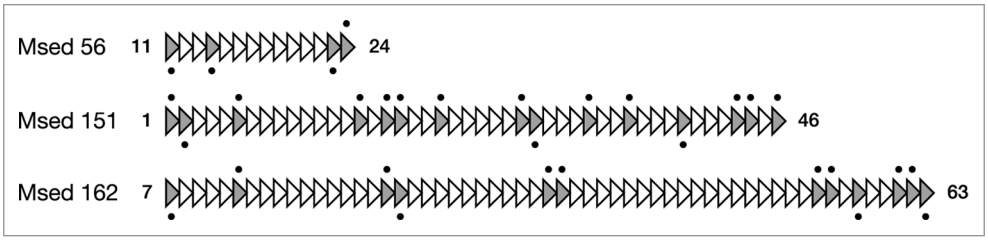

Figure 2. In silico determination of multiple spacer matches to the bicaudavirus ATV in three CRISPR loci of the crenarchaeon *M. sedula*. Repeat-spacer units from sections of type I-A_1 CRISPR arrays are depicted as arrowheads, directed away from the leader. Numbers to the left and right delineate the range of repeat-spacer units depicted. Shaded units yield close matches (≤ 5 nt mismatches) to protospacers in ATV. The orientation of the matching spacers with respect to the ATV genome is indicated by a dot above or below the shaded arrowheads.

was equal to GG in reducing transformation efficiency, consistent again with a reduced level of PAM recognition specificity operating during interference.

Almendros et al.[28] have taken this a stage further by demonstrating clear differences between the acquisition and interference protospacer motifs for a constitutive type I-F system in *E. coli*. They showed that a C located either at PAM position -2 or immediately upstream of the PAM (at position -3) resulted in interference, while the presence of both C's produced enhanced interference effects (**Table 1**).

The CRISPR system that diverges from this emerging consensus is the genetically manipulated type I-E system of *E. coli* where the PAM triplet, or more precisely the -2T and -3A positions, since the -1G position lies within the protospacer in this system,[4,13] appears to be critical for interference. Three of six possible single nucleotide mutations of the A and T were shown to produce strongly reduced interference.[23] Semenova et al.[23] demonstrated further that in this system, maintenance of seven of the first eight adjacent base pairs of the crRNA-protospacer hybrid (a "seed" sequence) were essential for effective interference. This highly specific interaction of the type I-E interference complex may also explain why the transformants generally evade interference via point mutations in the protospacer or PAM sequence, whereas in several other type I and type II systems and a type III-B system, elimination of matching CRISPR spacers or loss or mutation of Cas or Cmr proteins, are a much more common response to interference.[24,26,27,57-60]

Structural studies on the *E. coli* type I-E system provided evidence for type I-E-specific CasA protein, a predicted Cas8 homolog, interacting via a loop region with the PAM sequence located on the DNA strand complementary to the crRNA rather than to dsDNA or the non-targeted DNA strand.[29] Furthermore, the finding that the PAM sequence was important for initial binding of the interference complex to the protospacer, and not the targeted DNA strand, led to the proposal that the PAM sequence facilitates protospacer recognition.[29] In contrast, in a parallel study of the type II system of *Streptococcus pyogenes*, Jinek et al.[61] demonstrated that the PAM sequence was recognized exclusively on the opposite, non-complementary DNA strand during interference, consistent with the operation of fundamentally different molecular interference mechanisms in the type I-E and type II systems.

PAM Sequences and Type III Systems

The best characterized type III interference systems are the type III-A Csm DNA-targeting system of *Staphylococcus epidermidis*[62,63] and a type III-B RNA targeting system of the archaeon *Pyrococcus furiosus*.[20,64] No PAM-dependent spacer acquisition data are available for these organisms but experimental evidence suggests that specific PAM sequences are not recognized during interference. For the type III-A Csm system of *S. epidermidis*, evidence was presented that mismatched base pairing between the 5'-tag of the crRNA and the PAM region was sufficient to ensure interference.[63] Moreover, an antisense CRISPR RNA targeted by the type III-B Cmr system of *P. furiosus* was cleaved despite perfect matching of the 5'-tag of the crRNA to the antisense RNA substrate.[20] Similar evidence for PAM-independent interference was obtained for another RNA-targeting type III-B Cmr system of *S. solfataricus*,[65] and for a different type of III-B Cmr system of *Sulfolobus islandicus* putatively implicated in transcription-dependent DNA targeting.[60]

Thus, there appears to be no dependence of type III interference on PAM sequences. Indeed, among 126 available archaeal genome sequences (www.ebi.ac.uk/genomes/archaea.html), from early 2012, a total of 89 type III systems were represented, 51 of which are present as stand-alone gene cassettes and only 13 were linked exclusively to acquisition gene cassettes.[45] Therefore, these independent modules must function by utilizing spacers accumulated by type I acquisition systems in archaea, or by type I or type II acquisition systems in bacteria.[60] Consistent with this inference, Deng et al.[60] have demonstrated experimentally that a type III-B CRISPR interference module of *S. islandicus* can share crRNA processing Cas6 and specific spacers with co-existing type I systems.[60]

In summary, this widespread chromosomal uncoupling of type III interference modules from acquisition modules, as well as their lack of dependence on PAM sequences, seems to be a precondition for the occurrence of interference module exchange between organisms and the functional coupling to non-cognate CRISPR loci as well as to other types of acquisition modules.[4]

Intracellular Co-existence of Similar CRISPR Types Recognizing Different PAMs

Organisms often carry different CRISPR-based systems. For example, *S. thermophilus*

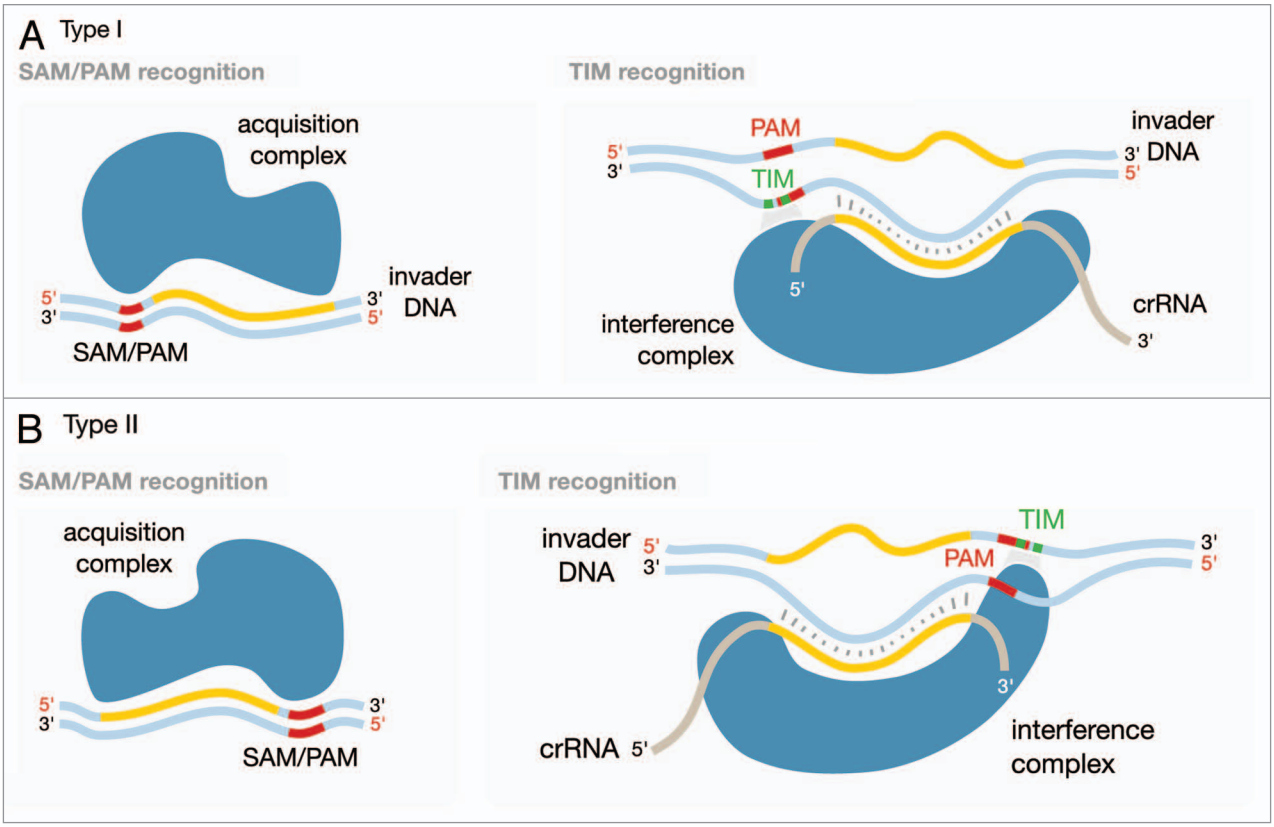

Figure 3. Overview of putative PAM, SAM and TIM interactions during acquisition and interference in type I and type II CRISPR systems. (**A**) The spacer acquisition motif (SAM) is recognized on the invader DNA by the Cas protein acquisition complex, which leads to the protospacer being excised by a putative ruler mechanism[14] and reinserted into a CRISPR locus by another putative ruler mechanism.[42] During interference by type I systems the target interference motif (TIM), on the crRNA-complementary DNA strand, is recognized by the Cas protein-crRNA complex where both TIM recognition and crRNA annealing are required for successful invader cleavage.[29] (**B**) In type II systems, the SAM/PAM motif is inferred to be recognized by a mechanism related to the type I system but inverted on the dsDNA whereas TIM recognition occurs on the non-complementary DNA strand to the crRNA.

contains types I, II and III-A systems,[10,25] and many thermophilic archaea exhibit type I and type III systems.[4] The diverse type III systems are likely to provide the host with a variety of interference options, often by sharing CRISPR loci of type I and II systems, and sometimes utilizing their CRISPR RNA processing enzymes.[41,60,64] There is also evidence of co-functionality of different subfamilies of type I interference complexes with different PAM sequences. For example, a subfamily I-A_1-specific interference protein Cas7, encoded adjacent to CRISPR loci C and D (PAM - CCN) of *S. solfataricus*, was found to be complexed with crRNAs from both subfamily I-A_1 CRISPR loci and from subfamily I-A_2 CRISPR loci A and B (PAM-TCN).[66] Moreover, a subfamily I-A_1 interference complex was shown to target protospacers with the subfamily I-A_2 PAM.[24] These observations further reinforce the view that PAM sequence recognition during interference differs from that occurring in the acquisition step.

Such flexibility of PAM sequence recognition during interference potentially renders the immune systems more versatile in that invading genetic elements will be unable to avoid targeting by incurring, for example, a single nucleotide mutation in a PAM sequence. Furthermore, this versatility can be extended for an organism by accumulating diverse interference modules, especially those of type III, which appear to exhibit a range of different targeting mechanisms, some of which remain to be elucidated.[20,60,64,65,67]

Conclusions and Perspectives

Both spacer acquisition and interference are dependent to some extent on PAM sequences. Currently, we know little about the molecular mechanism of PAM recognition during spacer acquisition, although preliminary evidence suggests that an initial, specific cleavage occurs downstream from the PAM in the type I-A system of *Sulfolobus* with a secondary cut directed by a ruler mechanism.[14] In contrast, during interference, the PAM sequence is recognized on opposite DNA strands for type I-E and type II-A systems.[29,61] Since spacer acquisition and protospacer interference must utilize fundamentally different molecular mechanisms, as envisioned in **Figure 3**, we consider that features of the PAM sequence recognized in these processes should be defined separately. Our proposal is to retain the acronym PAM for the conserved signature sequence and we prefer protospacer associated motif to the originally proposed protospacer adjacent motif[21] because some PAMs have recently been shown to include protospacer nucleotides.[12,13,42,47] Further, we suggest using the acronyms SAM for spacer acquisition motif and TIM for the target interference

motif. These motifs are defined further, and justified separately, below.

Protospacer associated motif (PAM). Consensus conserved sequence motif occurring at one end of predicted protospacers for each CRISPR-based system. PAMs are identified by sequence alignments of genetic elements containing sequences that match spacers within individual CRISPR arrays. The identity of the consensus PAM sequence probably depends on two molecular processes: (1) motif recognition by the spacer acquisition protein complex and (2) the subsequent selection of protospacers for targeting (i.e., those targeted more efficiently probably due to their exhibiting optimal interference motifs).

Spacer acquisition motif (SAM). Functional motif associated with a protospacer and recognized by the spacer acquisition machinery of each CRISPR-based system prior to protospacer excision. At present, the mode of recognition of the PAM sequence, and the DNA strand(s), remain unknown. Multiple SAMs may occur for a given PAM but the predominant or consensus SAM is likely to match the PAM but may be DNA-strand specific.

Target interference motif (TIM). Functional motifs associated with a protospacer and recognized by the DNA interference complex for each type I and type II CRISPR-Cas system. Multiple TIMs can occur for a single PAM as has been demonstrated experimentally for type I-A, I-B, I-E and type II-A systems, and the sequences are strand specific.[29,61]

In conclusion, we have used the order 5'-PAM-protospacer throughout this article for CRISPR type I systems, and protospacer-PAM-3' for type II systems for defining the PAM sequence. In a sense, the strand selection for the PAM sequence is arbitrary, and both orientations are widely used.[18,21,68] However, this lack of strand specificity is unlikely to apply to SAM and does not apply to TIM, for which the recognized motif is located on the crRNA complementary DNA strand, for a type I-E system[29] and on the non-complementary strand for a type II-A system[60] (**Fig. 3**).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Ling Deng contributed with helpful and insightful discussions. The research was supported by grants from the Danish Natural Science Research Council, Copenhagen University and the Spanish Ministerio de Ciencia e Innovación (BIO2011-24417).

References

1. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. Science 2010; 327:167-70; PMID:20056882; http://dx.doi.org/10.1126/science.1179555.
2. Terns MP, Terns RM. CRISPR-based adaptive immune systems. Curr Opin Microbiol 2011; 14:321-7; PMID:21531607; http://dx.doi.org/10.1016/j.mib.2011.03.005.
3. Makarova KS, Haft DH, Barrangou R, Brouns SJJ, 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.
4. 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.
5. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. Nature 2012; 482:331-8; PMID:22337052; http://dx.doi.org/10.1038/nature10886.
6. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol 2005; 60:174-82; PMID:15791728; http://dx.doi.org/10.1007/s00239-004-0046-3.
7. Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 2005; 151:653-63; PMID:15758212; http://dx.doi.org/10.1099/mic.0.27437-0.
8. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 2005; 151:2551-61; PMID:16079334; http://dx.doi.org/10.1099/mic.0.28048-0.
9. Horvath P, Romero DA, Coûté-Monvoisin AC, Richards M, Deveau H, Moineau S, et al. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J Bacteriol 2008; 190:1401-12; PMID:18065539; http://dx.doi.org/10.1128/JB.01415-07.
10. 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.
11. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol 2008; 190:1390-400; PMID:18065545; http://dx.doi.org/10.1128/JB.01412-07.
12. Yosef I, Goren MG, Qimron U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. Nucleic Acids Res 2012; 40:5569-76; PMID:22402487; http://dx.doi.org/10.1093/nar/gks216.
13. Swarts DC, Mosterd C, van Passel MWJ, Brouns SJJ. CRISPR interference directs strand specific spacer acquisition. PLoS One 2012; 7:e35888; PMID:22558257; http://dx.doi.org/10.1371/journal.pone.0035888.
14. Erdmann S, Garrett RA. Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. Mol Microbiol 2012; 85:1044-56; PMID:22834906; http://dx.doi.org/10.1111/j.1365-2958.2012.08171.x.
15. Tang TH, Bachellerie JP, Rozhdestvensky T, Bortolin ML, Huber H, Drungowski M, et al. Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. Proc Natl Acad Sci USA 2002; 99:7536-41; PMID:12032318; http://dx.doi.org/10.1073/pnas.112047299.
16. Tang TH, Polacek N, Zywicki M, Huber H, Brügger K, Garrett R, et al. Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. Mol Microbiol 2005; 55:469-81; PMID:15659164; http://dx.doi.org/10.1111/j.1365-2958.2004.04428.x.
17. Lillestøl RK, Redder P, Garrett RA, Brügger K. A putative viral defence mechanism in archaeal cells. Archaea 2006; 2:59-72; PMID:16877322; http://dx.doi.org/10.1155/2006/542818.
18. Lillestøl RK, Shah SA, Brügger K, Redder P, Phan H, Christiansen J, et al. CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. Mol Microbiol 2009; 72:259-72; PMID:19239620; http://dx.doi.org/10.1111/j.1365-2958.2009.06641.x.
19. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 2008; 321:960-4; PMID:18703739; http://dx.doi.org/10.1126/science.1159689.
20. Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, et al. Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. Mol Cell 2012; 45:292-302; PMID:22227116; http://dx.doi.org/10.1016/j.molcel.2011.10.023.
21. Mojica FJM, 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.
22. Shah SA, Hansen NR, Garrett RA. Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. Biochem Soc Trans 2009; 37:23-8; PMID:19143596; http://dx.doi.org/10.1042/BST0370023.
23. 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 USA 2011; 108:10098-103; PMID:21646539; http://dx.doi.org/10.1073/pnas.1104144108.
24. Gudbergsdottir S, Deng L, Chen Z, Jensen JVK, 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.
25. Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. Nucleic Acids Res 2011; 39:9275-82; PMID:21813460; http://dx.doi.org/10.1093/nar/gkr606.

26. 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.
27. Lopez-Sanchez MJ, Sauvage E, Da Cunha V, Clermont D, Ratsima Hariniaina E, Gonzalez-Zorn B, et al. The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome. Mol Microbiol 2012; 85:1057-71; PMID:22834929; http://dx.doi.org/10.1111/j.1365-2958.2012.08172.x.
28. Almendros C, Guzmán NM, Díez-Villaseñor C, García-Martínez J, Mojica FJM. Target motifs affecting natural immunity by a constitutive CRISPR-Cas system in *Escherichia coli*. PLoS One 2012; 7:e50797; PMID:23189210; http://dx.doi.org/10.1371/journal.pone.0050797.
29. Sashital DG, Wiedenheft B, Doudna JA. Mechanism of foreign DNA selection in a bacterial adaptive immune system. Mol Cell 2012; 46:606-15; PMID:22521690; http://dx.doi.org/10.1016/j.molcel.2012.03.020.
30. Manica A, Zebec Z, Teichmann D, Schleper C. *In vivo* activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. Mol Microbiol 2011; 80:481-91; PMID:21385233; http://dx.doi.org/10.1111/j.1365-2958.2011.07586.x.
31. Wiedenheft B, van Duijn E, Bultema JB, Waghmare SP, Zhou K, Barendregt A, et al. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. Proc Natl Acad Sci USA 2011; 108:10092-7; PMID:21536913; http://dx.doi.org/10.1073/pnas.1102716108.
32. Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun 2012; 3:945; PMID:22781758; http://dx.doi.org/10.1038/ncomms1937.
33. Westra ER, van Erp PB, Künne T, Wong SP, Staals RH, Seegers CL, et al. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol Cell 2012; 46:595-605; PMID:22521689; http://dx.doi.org/10.1016/j.molcel.2012.03.018.
34. Garrett RA, Prangishvili D, Shah SA, Reuter M, Stetter KO, Peng X. Metagenomic analyses of novel viruses and plasmids from a cultured environmental sample of hyperthermophilic neutrophiles. Environ Microbiol 2010; 12:2918-30; PMID:20545752; http://dx.doi.org/10.1111/j.1462-2920.2010.02266.x.
35. Kunin V, Sorek R, Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol 2007; 8:R61; PMID:17442114; http://dx.doi.org/10.1186/gb-2007-8-4-r61.
36. Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W, Doudna JA. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. Structure 2009; 17:904-12; PMID:19523907; http://dx.doi.org/10.1016/j.str.2009.03.019.
37. Babu M, Beloglazova N, Flick R, Graham C, Skarina T, Nocek B, et al. A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Mol Microbiol 2011; 79:484-502; PMID:21219465; http://dx.doi.org/10.1111/j.1365-2958.2010.07465.x.
38. Nam KH, Ding F, Haitjema C, Huang Q, DeLisa MP, Ke A. Double-stranded endonuclease activity in *Bacillus halodurans* clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. J Biol Chem 2012; 287:35943-52; PMID:22942283; http://dx.doi.org/10.1074/jbc.M112.382598.
39. Beloglazova N, Brown G, Zimmerman MD, Proudfoot M, Makarova KS, Kudritska M, et al. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. J Biol Chem 2008; 283:20361-71; PMID:18482976; http://dx.doi.org/10.1074/jbc.M803225200.
40. Makarova KS, Anantharaman V, Aravind L, Koonin EV. Live virus-free or die: coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes. Biol Direct 2012; 7:40; PMID:23151069; http://dx.doi.org/10.1186/1745-6150-7-40.
41. Zhang J, Kasciukovic T, White MF. The CRISPR associated protein Cas4 Is a 5' to 3' DNA exonuclease with an iron-sulfur cluster. PLoS One 2012; 7:e47232; PMID:23056615; http://dx.doi.org/10.1371/journal.pone.0047232.
42. Díez-Villaseñor C, Guzmán NM, Almendros C, García-Martínez J, Mojica FJM. CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of *Escherichia coli*. RNA Biol, In press.
43. Westra ER, Brouns SJ. The rise and fall of CRISPRs--dynamics of spacer acquisition and loss. Mol Microbiol 2012; 85:1021-5; PMID:22804962; http://dx.doi.org/10.1111/j.1365-2958.2012.08170.x.
44. Pul U, Wurm R, Arslan Z, Geissen R, Hofmann N, Wagner R. Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. Mol Microbiol 2010; 75:1495-512; PMID:20132443; http://dx.doi.org/10.1111/j.1365-2958.2010.07073.x.
45. Westra ER, Pul U, Heidrich N, Jore MM, Lundgren M, Stratmann T, et al. H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. Mol Microbiol 2010; 77:1380-93; PMID:20659289; http://dx.doi.org/10.1111/j.1365-2958.2010.07315.x.
46. Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, et al. Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. Mol Microbiol 2010; 77:1367-79; PMID:20624226; http://dx.doi.org/10.1111/j.1365-2958.2010.07265.x.
47. Goren MG, Yosef I, Auster O, Qimron U. Experimental definition of a clustered regularly interspaced short palindromic duplicon in *Escherichia coli*. J Mol Biol 2012; 423:14-6; PMID:22771574; http://dx.doi.org/10.1016/j.jmb.2012.06.037.
48. Shah SA, Garrett RA. CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. Res Microbiol 2011; 162:27-38; PMID:20863886; http://dx.doi.org/10.1016/j.resmic.2010.09.001.
49. Charbonnier F, Forterre P. Comparison of plasmid DNA topology among mesophilic and thermophilic eubacteria and archaebacteria. J Bacteriol 1994; 176:1251-9; PMID:8113163.
50. van der Ploeg JR. Analysis of CRISPR in *Streptococcus mutans* suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. Microbiology 2009; 155:1966-76; PMID:19383692; http://dx.doi.org/10.1099/mic.0.027508-0.
51. Richter C, Gristwood T, Clulow JS, Fineran PC. *In vivo* protein interactions and complex formation in the *Pectobacterium atrosepticum* subtype I-F CRISPR/Cas System. PLoS One 2012; 7:e49549; PMID:23226499; http://dx.doi.org/10.1371/journal.pone.0049549.
52. Díez-Villaseñor C, Almendros C, García-Martínez J, Mojica FJM. Diversity of CRISPR loci in *Escherichia coli*. Microbiology 2010; 156:1351-61; PMID:20133361; http://dx.doi.org/10.1099/mic.0.036046-0.
53. Guo L, Brügger K, Liu C, Shah SA, Zheng H, Zhu Y, et al. Genome analyses of Icelandic strains of *Sulfolobus islandicus*, model organisms for genetic and virus-host interaction studies. J Bacteriol 2011; 193:1672-80; PMID:21278296; http://dx.doi.org/10.1128/JB.01487-10.
54. Häring M, Vestergaard G, Rachel R, Chen L, Garrett RA, Prangishvili D. Virology: independent virus development outside a host. Nature 2005; 436:1101-2; PMID:16121167; http://dx.doi.org/10.1038/4361101a.
55. 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.
56. Garcia-Heredia I, Martin-Cuadrado AB, Mojica FJ, Santos F, Mira A, Antón J, et al. Reconstructing viral genomes from the environment using fosmid clones: the case of haloviruses. PLoS One 2012; 7:e33802; PMID:22479446; http://dx.doi.org/10.1371/journal.pone.0033802.
57. Tyson GW, Banfield JF. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ Microbiol 2008; 10:200-7; PMID:17894817.
58. 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.
59. Deng L, Kenchappa CS, Peng X, She Q, Garrett RA. Modulation of CRISPR locus transcription by the repeat-binding protein Cbp1 in *Sulfolobus*. Nucleic Acids Res 2012; 40:2470-80; PMID:22139923; http://dx.doi.org/10.1093/nar/gkr1111.
60. Deng L, Garrett RA, Shah SA, Peng X, She Q. A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. Mol Microbiol 2013; In press; PMID:23320564; http://dx.doi.org/10.1111/mmi.12152.
61. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012; 337:816-21; PMID:22745249; http://dx.doi.org/10.1126/science.1225829.
62. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in *staphylococci* by targeting DNA. Science 2008; 322:1843-5; PMID:19095942; http://dx.doi.org/10.1126/science.1165771.
63. 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.
64. Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, et al. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell 2009; 139:945-56; PMID:19945378; http://dx.doi.org/10.1016/j.cell.2009.07.040.
65. Zhang J, Rouillon C, Kerou M, Reeks J, Brügger K, Graham S, et al. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. Mol Cell 2012b; 45:303-13; PMID:22227115; http://dx.doi.org/10.1016/j.molcel.2011.12.013.
66. Lintner NG, Kerou M, Brumfield SK, Graham S, Liu H, Naismith JH, et al. Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). J Biol Chem 2011; 286:21643-56; PMID:21507944; http://dx.doi.org/10.1074/jbc.M111.238485.
67. Hatoum-Aslan A, Maniv I, Marraffini LA. Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. Proc Natl Acad Sci USA 2011; 108:21218-22; PMID:22160698; http://dx.doi.org/10.1073/pnas.1112832108.
68. Westra ER, Swarts DC, Staals RH, Jore MM, Brouns SJ, van der Oost J. The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. Annu Rev Genet 2012; 46:311-39; PMID:23145983; http://dx.doi.org/10.1146/annurev-genet-110711-155447.