

CRISPR/Cas, the Immune System of Bacteria and Archaea

Philippe Horvath^{1*} and Rodolphe Barrangou^{2*}

Microbes rely on diverse defense mechanisms that allow them to withstand viral predation and exposure to invading nucleic acid. In many Bacteria and most Archaea, clustered regularly interspaced short palindromic repeats (CRISPR) form peculiar genetic loci, which provide acquired immunity against viruses and plasmids by targeting nucleic acid in a sequence-specific manner. These hypervariable loci take up genetic material from invasive elements and build up inheritable DNA-encoded immunity over time. Conversely, viruses have devised mutational escape strategies that allow them to circumvent the CRISPR/Cas system, albeit at a cost. CRISPR features may be exploited for typing purposes, epidemiological studies, host-virus ecological surveys, building specific immunity against undesirable genetic elements, and enhancing viral resistance in domesticated microbes.

Microbes have devised various strategies that allow them to survive exposure to foreign genetic elements. Although outpopulated and preyed upon by abundant and ubiquitous viruses, microbes routinely survive, persist, and occasionally thrive in hostile and competitive environments. The constant exposure to exogenous DNA via transduction, conjugation, and transformation have forced microbes to establish an array of defense mechanisms that allow the cell to recognize and distinguish incoming “foreign” DNA, from “self” DNA and to survive exposure to invasive elements. These systems maintain genetic integrity, yet occasionally allow exogenous DNA uptake and conservation of genetic material advantageous for adaptation to the environment. Certain strategies, such as prevention of adsorption, blocking of injection, and abortive infection, are effective against viruses; other defense systems specifically target invading nucleic acid, such as the restriction-modification system (R-M) and the use of sugar-nonspecific nucleases. Recently, an adaptive microbial immune system, clustered regularly interspaced short palindromic repeats (CRISPR) has been identified that provides acquired immunity against viruses and plasmids.

CRISPR represents a family of DNA repeats found in most archaeal (~90%) and bacterial (~40%) genomes (1–3). Although the initial discovery of a CRISPR structure was made fortuitously in *Escherichia coli* in 1987, the acronym was coined in 2002, after similar structures were observed in genomes of various Bacteria and Archaea (1). CRISPR loci typically consist of several noncontiguous direct repeats separated by stretches of variable sequences called spacers (which mostly correspond to segments of captured

viral and plasmid sequences) and are often adjacent to *cas* genes (CRISPR-associated) (Fig. 1). *cas* genes encode a large and heterogeneous

family of proteins that carry functional domains typical of nucleases, helicases, polymerases, and polynucleotide-binding proteins (4). CRISPR, in combination with Cas proteins, forms the CRISPR/Cas systems. Six “core” *cas* genes have been identified, including the universal markers of CRISPR/Cas systems *cas1* (COG1518) and *cas2* (COG1343, COG3512, occasionally in a fused form with other *cas* genes). Besides the *cas1* to *cas6* core genes, subtype-specific genes and genes encoding “repeat-associated mysterious proteins” (RAMP) have been identified and grouped into subtypes functionally paired with particular CRISPR repeat sequences (4–8). The size of CRISPR repeats and spacers varies between 23 to 47 base pairs (bp) and 21 to 72 bp, respectively. Generally, CRISPR repeat sequences are highly conserved within a given CRISPR locus, but a large assortment of repeat sequences has been shown across microbial species (1, 9). Most repeat sequences are partially palindromic, having the potential to form stable, highly conserved secondary structures (7). The number

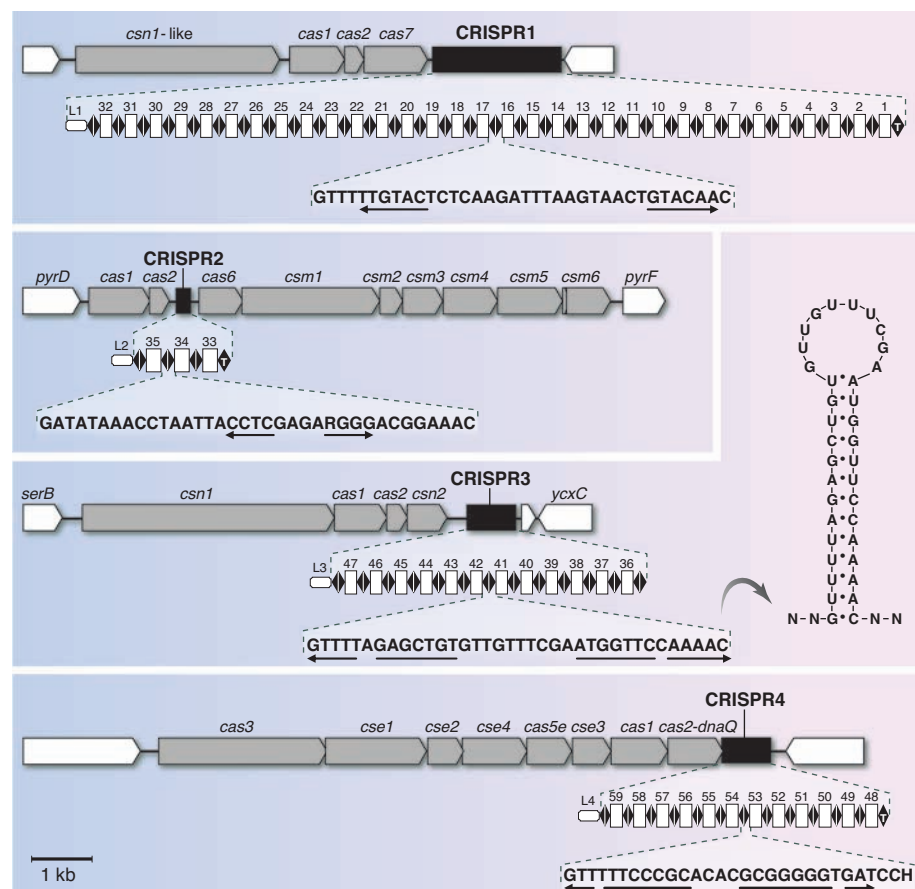


Fig. 1. Overview of the four CRISPR/*cas* systems present in *Streptococcus thermophilus* DGCC7710. For each system, gene organization is depicted on the top, with *cas* genes in gray, and the repeat-spacer array in black. Below the gene scheme, the repeat and spacer (captured phage or plasmid nucleic acid) content is detailed as black diamonds (T, terminal repeat) and white rectangles, respectively. Bottom line, consensus repeat sequence. L1 to L4, leader sequences. The predicted secondary structure of the CRISPR3 repeat is shown on the right. *S. thermophilus* CRISPR2, CRISPR3, and CRISPR4 systems are homologous to the CRISPR systems of *Staphylococcus epidermidis* (20), *Streptococcus mutans* (19), and *E. coli* (28), respectively.

¹Danisco France SAS, BP10, F-86220 Dange-Saint-Romain, France. ²Danisco USA, Inc., 3329 Agriculture Drive, Madison, WI 53716, USA.

*To whom correspondence should be addressed. E-mail: philippe.horvath@danisco.com (P.H.), rodolphe.barrangou@danisco.com (R.B.)

of repeat-spacer units is documented to reach 375 (*Chloroflexus* sp. Y-400-fl), but most loci commonly contain fewer than 50 units, as exemplified in lactic acid bacteria genomes (8). Microbes may contain more than one CRISPR locus; up to 18 such loci have been identified in *Methanocaldococcus jannaschii*, totaling more than 1% of the genome (10). CRISPRs are typically located on the chromosome, although some have been identified on plasmids (11–13).

The CRISPR loci have highly diverse and hypervariable spacer sequences, even between closely related strains (14–16), which were initially exploited for typing purposes. A variety of putative roles for CRISPR sequences was originally suggested, including chromosomal rearrangement, modulation of expression of neighboring genes, target for DNA binding proteins, replicon partitioning, and DNA repair (5). In 2005, three independent in silico studies reported homology between spacer sequences and extrachromosomal elements, such as viruses and plasmids (11, 14, 15). This led to the hypothesis that CRISPR may provide adaptive immunity against foreign genetic elements (6).

A Vast Spectrum of Immunity

In 2007, it was shown in *Streptococcus thermophilus* that during natural generation of phage-resistant variants, bacteria commonly alter their CRISPR loci by polarized (i.e., at the leader end) incorporation of CRISPR repeat-spacer units (Fig. 1) (17, 18), consistent with observed spacer hyper-variability at the leader end of CRISPR loci in various strains (14, 16). The integrated sequences were identical to those of the phages used in the challenge, which suggested that they originate from viral nucleic acid. To determine whether CRISPR impacts phage resistance, spacer content was altered via genetic engineering, which showed that spacer addition can provide novel phage resistance, whereas spacer deletion could result in loss of phage resistance (17). These findings were confirmed in *Streptococcus mutans*, where phage-resistant mutants acquired novel CRISPR spacers with sequences matching the phage genome, in vitro and in vivo (19). Although the ubiquitous and predatory nature of phages may explain the overwhelming representation of phage sequences in CRISPR loci, CRISPR spacers can also interfere with both plasmid conjugation and transformation, as shown in *Staphylococcus epidermidis* (20). Furthermore, several metagenomic studies investigating host-virus population dynamics showed that CRISPR loci evolve in response to viral predation and that CRISPR spacer content and sequential order provide insights both historically and geographically (21–24).

The ability to provide defense against invading genetic elements seems to render CRISPR/Cas systems particularly desirable in hostile environments and may explain their propensity to be transferred horizontally between sometimes distant organisms (12). There is extensive evidence that defense systems such as CRISPR have

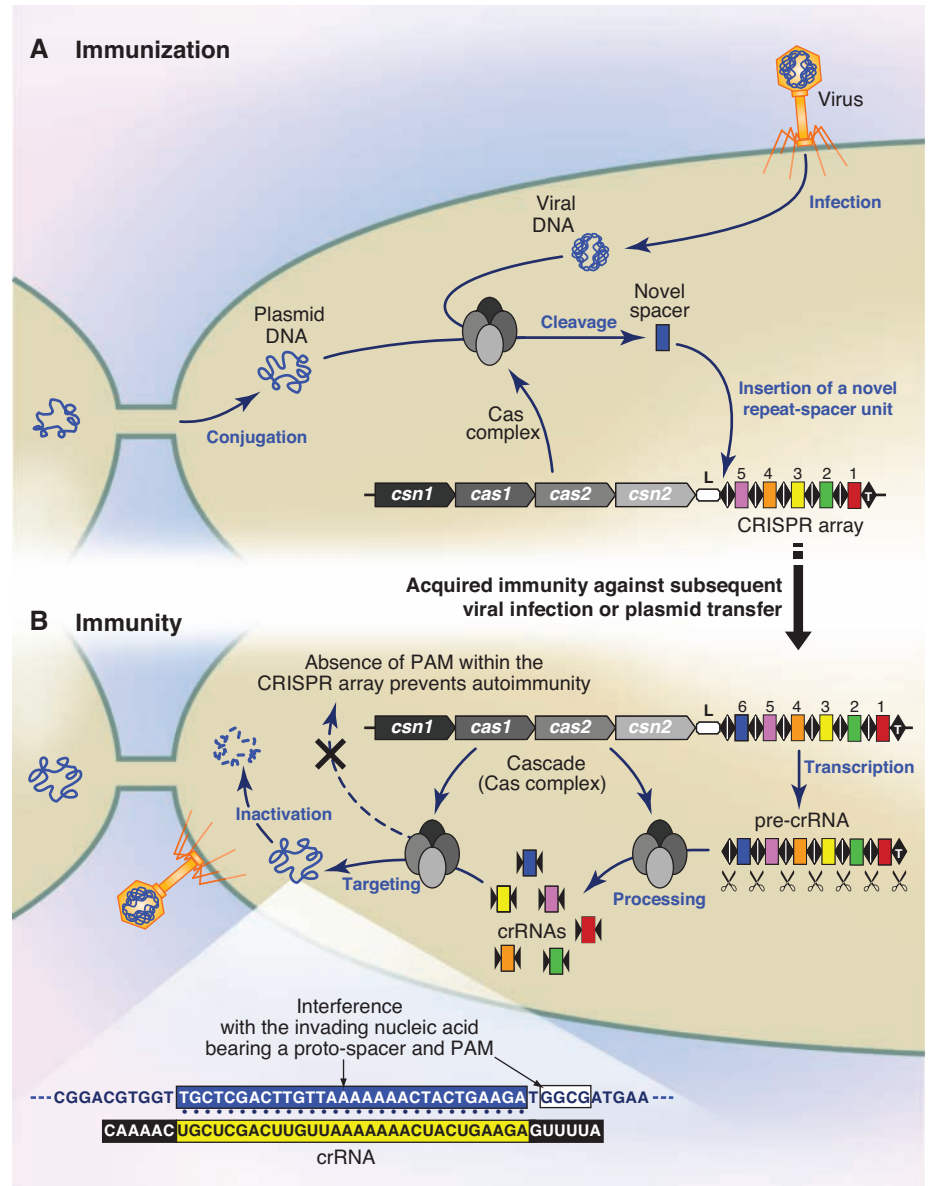


Fig. 2. Overview of the CRISPR/Cas mechanism of action. (A) Immunization process: After insertion of exogenous DNA from viruses or plasmids, a Cas complex recognizes foreign DNA and integrates a novel repeat-spacer unit at the leader end of the CRISPR locus. (B) Immunity process: The CRISPR repeat-spacer array is transcribed into a pre-crRNA that is processed into mature crRNAs, which are subsequently used as a guide by a Cas complex to interfere with the corresponding invading nucleic acid. Repeats are represented as diamonds, spacers as rectangles, and the CRISPR leader is labeled L.

undergone horizontal transfer between genomes, notably differences observed in codon bias, GC content variability, their presence on mobile genetic elements, the presence of neighboring insertion sequence elements, and their variable presence and location in closely related genomes. This is in agreement with the lack of congruence between the phylogenetic relation of various CRISPR elements and that of the organisms in which they are found (8, 12). This horizontal gene transfer may be mediated by plasmids, megaplasmids, and even prophages, all of which are documented to carry CRISPR loci (2).

Given the variety of defense systems in microbes and their role in controlling the presence

of plasmids, prophages, transposons, and, perhaps, chromosomal sequences, studies should investigate whether CRISPR/Cas systems preferentially target certain elements and could determine whether they are symbiotic or mutually exclusive with other defense systems.

Idiosyncrasies of the CRISPR/Cas Mechanism of Action

The mechanism by which CRISPR provides resistance against foreign genetic elements is not fully characterized (Fig. 2). Even so, the functional link between Cas and CRISPR repeats has been inferred from the congruence observed between their sequence patterns. *cas* genes provide

CRISPR-encoded immunity, because inactivating the CRISPR1-associated *cas7* gene (Fig. 1) impairs the ability of the host to integrate novel CRISPR spacers after phage exposure (17), which suggests that it is necessary for recognizing foreign nucleic acid and/or integrating the novel repeat-spacer unit. Cas1 appears to be a double-stranded DNA (dsDNA) endonuclease involved in the immunization process (25). It has also been proposed that Cas2 may act as a sequence-specific endoribonuclease that cleaves uracil-rich single-stranded RNAs (ssRNAs) (26). The mechanistic steps involved in invasive element recognition, novel repeat manufacturing, and spacer selection and integration into the CRISPR locus remain uncharacterized.

Although some Cas proteins are involved in the acquisition of novel spacers, others provide CRISPR-encoded phage resistance and interfere with invasive genetic elements. Mechanistically, although defense is spacer-encoded, the information that lies within the CRISPR repeat-spacer array becomes available to the Cas machinery through transcription. The CRISPR leader, defined as a low-complexity, A/T-rich, noncoding sequence, located immediately upstream of the first repeat, likely acts as a promoter for the transcription of the repeat-spacer array into a CRISPR transcript, the pre-crRNA (13, 27). The full-length pre-crRNA is subsequently processed into specific small RNA molecules that correspond to a spacer flanked by two partial repeats (27–29). In *E. coli*, processing is achieved by a multimeric complex of Cas proteins named Cascade (CRISPR-associated complex for antiviral defense), which specifically cleaves the pre-crRNA transcript within the repeat sequence to generate small CRISPR RNAs, crRNAs (28). Similarly, in *Pyrococcus*, Cas6 is an endoribonuclease that cleaves the pre-crRNA transcript into crRNA units that include a partial [8-nucleotide (nt)] repeat sequence at the 5' end, as part of the Cas-crRNA complex (27, 29, 30). The crRNAs seem to specifically guide the Cas interference machinery toward foreign nucleic acid molecules that match its sequence, which leads ultimately to degradation of the invading element (30). The involvement of *cas* genes in CRISPR defense was originally demonstrated when inactivating the CRISPR1-associated *csn1*-like gene (Fig. 1) resulted in loss of phage resistance despite the presence of matching spacers (17).

The observation that CRISPR spacers match both sense and antisense viral DNA led to the hypothesis that some CRISPR/Cas systems may target dsDNA, and this was confirmed by disruption of target DNA with an intron (the excision of which restores the

native mRNA) on a plasmid that allows conjugation despite the presence of a matching CRISPR spacer (20). Conversely, the *Pyrococcus* CRISPR effector complex, a ribonucleoprotein complex that consists of crRNA and Cas proteins, targets invader RNA by complementary-dependent cleavage, *in vitro* (30). Given the large diversity of CRISPR/Cas systems in Bacteria and Archaea (4, 6), it is likely that both DNA and RNA may be targets. More information is needed to establish and understand what the functional differences are among distinct CRISPR/Cas systems.

The initial hypothesis that CRISPR may mediate microbial immunity via RNA interference (RNAi) (6) is misguided. RNAi allows eukaryotic organisms exposed to foreign genetic material to silence the invading nucleic acid sequence before or after it integrates into the host chromosome, and/or to subvert cellular processes through a small interfering RNA guide (31). A key difference between RNAi and CRISPR-encoded immunity lies in the enzymatic machinery involved. Although both are mediated by a guide RNA in an inhibitory ribonucleoprotein complex, only Dicer, Slicer, and the RNA-induced silencing complex (RISC) may have analogous counterparts (6, 30). Mechanistically, although the short RNA duplexes at the core of RNAi are typically 21 to 28 nt in length (32), crRNAs are larger, because they contain a CRISPR spacer (23 to 47 nt) flanked by partial repeats. Also, RNA-dependent transcription generating dsRNA and using the cleaved target RNA seen in RNAi have not been characterized in the CRISPR/Cas systems. In other ways, the sequence-specific and adaptive CRISPR/Cas systems share sim-

ilarities with the vertebrate adaptive immune system, although CRISPR spacers are DNA-encoded and can be inherited by the progeny.

Circumventing CRISPR-Based Immunity

Even though CRISPR can provide high levels of phage resistance, a relatively small proportion of viruses retain the ability to infect the “immunized” host. These viral particles have specifically mutated the proto-spacer (sequence within the invading nucleic acid that matches a CRISPR spacer), with a single point mutation that allows the viruses to overcome immunity, which indicates that the selective pressure imposed by CRISPR can rapidly drive mutation patterns in viruses (17, 18, 23). Analysis of phage sequences adjacent to proto-spacers revealed the presence of conserved sequences, called CRISPR motifs (13, 16, 18, 19, 33, 34), or proto-spacer adjacent motifs (PAMs) (35). Phages may also circumvent the CRISPR/Cas system by mutating the CRISPR motif (18), which indicates that it is involved in CRISPR-encoded immunity. Additionally, CRISPR motif mutation can result in loss of phage resistance despite the presence of a matching CRISPR spacer (34). The absence of this motif in the CRISPR locus likely allows the system to act on the invading target DNA specifically and precludes an “autoimmune” response on the host chromosome (Fig. 2). Such a motif may not be necessary in CRISPR/Cas systems targeting RNA. Although proto-spacers seem to be randomly located on phage genomes, a given CRISPR spacer may be acquired independently by different lineages. It is thus tempting to speculate that CRISPR motifs also play a key role in the selection of spacers.

These mutations may have an impact on the amino acid sequence, as either nonsynonymous mutations or premature stop codons that truncate the viral protein (18). In addition to mutations, phages may also circumvent CRISPR-encoded immunity via deletion of the target sequence (18, 21). This perhaps indicates a strong cost associated with circumventing the CRISPR/Cas systems. Alternative strategies that allow viruses to escape CRISPR, such as suppressors that could interfere with crRNAs biogenesis or Cas machinery remain uncovered. Defense tactics employed by viruses to circumvent the CRISPR/Cas systems are yet another critical difference between RNAi and CRISPR: Eukaryotic viruses may express inhibitors such as dsRNA-binding proteins that interfere with the RNA silencing machinery (32), which are yet to be identified in response to CRISPR, whereas microbial viruses specifically mutate or recombine (21) the sequence

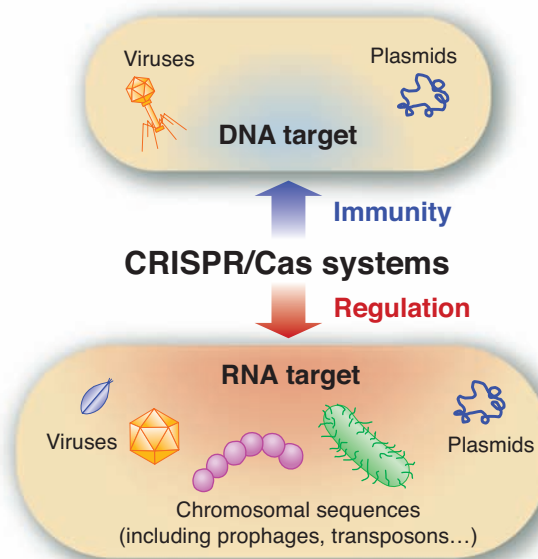


Fig. 3. CRISPR interference. The CRISPR/Cas systems may target either DNA or RNA to interfere with viruses, plasmids, prophages, or other chromosomally encoded sequences.

corresponding to the CRISPR spacer or that of the PAM.

The impact of CRISPR on phage genomes is illustrated by extensive genome recombination events observed in environmental phage populations in response to CRISPR (21). This contrasts with the fact that acquisition of novel CRISPR spacers does not seem to have a fitness cost for the host, apart from maintaining the CRISPR/Cas system as active.

Although it seems intuitive that CRISPR loci should not be able to expand indefinitely (21, 36), little is known about the parameters that define the optimal and maximum size of a CRISPR locus. Also, the fitness cost of CRISPR expansion in the host should be compared with that of CRISPR evasion in the virus populations, so as to determine whether prey or predators incur the higher evolutionary cost of this genetic warfare.

Although CRISPR loci primarily evolve via polarized addition of novel spacers at the leader end of the locus after phage exposure, internal spacer deletions have also been reported, likely occurring via homologous recombination between CRISPR repeats (1, 16, 18). Perhaps this allows the host to limit the expansion of the CRISPR locus so that the relative size of the locus does not increase to a detrimental level. The propensity of spacers located at the trailer end (opposite to the leader end) to be deleted preferentially would mitigate the loss of fitness associated with the deletion, because ancestral spacers would arguably provide resistance against viruses that were historically, but are not currently, present in the environment. The combination of locus expansion via spacer acquisition and contraction via spacer loss, in the context of rapid evolution in space and time because of viral predation, which generate a high level of spacer polymorphism, suggests that CRISPR loci undergo dynamic and rapid turnover on evolutionary time scales (16, 21, 36). Indeed, in microbes with an active system, CRISPR loci have been shown to be the most hypervariable genomic regions (21).

Applications and Future Directions

A priori, the concurrent presence of distinct defense systems against foreign genetic elements in Bacteria and Archaea seems inefficient and redundant, although it might reflect functional preferences and increased fitness. Because all

defense mechanisms have their advantages and caveats, the accumulation and combination of different systems would increase the selective pressure on invading elements and, consequently, could increase the chances of host survival by using multiple hurdles.

Because CRISPR spacers correspond to prior episodes of phage and plasmid exposure, they provide a historical and geographical—although limited—perspective as to the origin and paths of a particular strain, which may be used for ecological and epidemiological studies. Many intrinsic aspects of CRISPR-based immunity have provided avenues for industrial applications, including exploiting hypervariability for typing purposes, driving viral evolution, predicting and modulating virus resistance in domesticated microbes, and performing natural genetic tagging of proprietary strains. The inheritable nature of the CRISPR spacer content provides potential for perennial use of industrial microbes. Alternatively, the ability of CRISPR/Cas systems to impede the transfer of particular nucleic acid sequences (such as phage or plasmid DNA) into a host might be exploited via genetic engineering to specifically preclude the dissemination of undesirable genetic elements, such as antibiotic-resistance markers and genes harmful to humans and other living organisms. It may also be designed to limit the intracellular spread of mobile genetic elements such as insertion sequences and transposons. In addition to providing immunity, CRISPR/Cas systems that target RNA have the potential to affect the transcript stability of chromosomal elements (Fig. 3).

Although significant progress has been made in the last few years, many mechanistic aspects remain uncovered, notably vis-à-vis the immunization process (key elements involved in spacer selection and integration between repeats and/or possible involvement of degenerate infectious particles in building immunity) and the interference mechanism (other cellular components involved). Also, more knowledge is desirable regarding the elements necessary to have functional CRISPR/Cas systems and the basis for the absence of CRISPR in 60% of Bacteria.

References and Notes

1. R. Jansen, J. D. Embden, W. Gaastra, L. M. Schouls, *Mol. Microbiol.* **43**, 1565 (2002).
2. R. Sorek, V. Kunin, P. Hugenholtz, *Nat. Rev. Microbiol.* **6**, 181 (2008).

3. J. van der Oost, M. M. Jore, E. R. Westra, M. Lundgren, S. J. Brouns, *Trends Biochem. Sci.* **34**, 401 (2009).
4. D. H. Haft, J. Selengut, E. F. Mongodin, K. E. Nelson, *PLoS Comput. Biol.* **1**, e60 (2005).
5. K. S. Makarova, L. Aravind, N. V. Grishin, I. B. Rogozin, E. V. Koonin, *Nucleic Acids Res.* **30**, 482 (2002).
6. K. S. Makarova, N. V. Grishin, S. A. Shabalina, Y. I. Wolf, E. V. Koonin, *Biol. Direct* **1**, 7 (2006).
7. V. Kunin, R. Sorek, P. Hugenholtz, *Genome Biol.* **8**, R61 (2007).
8. P. Horvath *et al.*, *Int. J. Food Microbiol.* **131**, 62 (2009).
9. I. Grissa, G. Vergnaud, C. Pourcel, *BMC Bioinformatics* **8**, 172 (2007).
10. R. K. Lillestøl, P. Redder, R. A. Garrett, K. Brügger, *Archaea* **2**, 59 (2006).
11. F. J. Mojica, C. Díez-Villaseñor, J. García-Martínez, E. Soria, *J. Mol. Evol.* **60**, 174 (2005).
12. J. S. Godde, A. Bickerton, *J. Mol. Evol.* **62**, 718 (2006).
13. R. K. Lillestøl *et al.*, *Mol. Microbiol.* **72**, 259 (2009).
14. C. Pourcel, G. Salvignol, G. Vergnaud, *Microbiology* **151**, 653 (2005).
15. A. Bolotin, B. Quinquis, A. Sorokin, S. D. Ehrlich, *Microbiology* **151**, 2551 (2005).
16. P. Horvath *et al.*, *J. Bacteriol.* **190**, 1401 (2008).
17. R. Barrangou *et al.*, *Science* **315**, 1709 (2007).
18. H. Deveau *et al.*, *J. Bacteriol.* **190**, 1390 (2008).
19. J. R. van der Ploeg, *Microbiology* **155**, 1966 (2009).
20. L. A. Marraffini, E. J. Sontheimer, *Science* **322**, 1843 (2008).
21. A. F. Andersson, J. F. Banfield, *Science* **320**, 1047 (2008).
22. V. Kunin *et al.*, *Genome Res.* **18**, 293 (2008).
23. J. F. Heidelberg, W. C. Nelson, T. Schoenfeld, D. Bhaya, N. Ahmed, *PLoS ONE* **4**, e4169 (2009).
24. N. L. Held, R. J. Whitaker, *Environ. Microbiol.* **11**, 457 (2009).
25. B. Wiedenheft *et al.*, *Structure* **17**, 904 (2009).
26. N. Beloglazova *et al.*, *J. Biol. Chem.* **283**, 20361 (2008).
27. C. Hale, K. Kleppe, R. M. Terns, M. P. Terns, *RNA* **14**, 2572 (2008).
28. S. J. Brouns *et al.*, *Science* **321**, 960 (2008).
29. J. Carte, R. Wang, H. Li, R. M. Terns, M. P. Terns, *Genes Dev.* **22**, 3489 (2008).
30. C. R. Hale *et al.*, *Cell* **139**, 945 (2009).
31. C. C. Mello, D. Conte Jr., *Nature* **431**, 338 (2004).
32. G. Meister, T. Tuschl, *Nature* **431**, 343 (2004).
33. S. A. Shah, N. R. Hansen, R. A. Garrett, *Biochem. Soc. Trans.* **37**, 23 (2009).
34. E. Semenova, M. Nagornykh, M. Pyatnitskiy, I. I. Artamonova, K. Severinov, *FEMS Microbiol. Lett.* **296**, 110 (2009).
35. F. J. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, *Microbiology* **155**, 733 (2009).
36. G. W. Tyson, J. F. Banfield, *Environ. Microbiol.* **10**, 200 (2008).
37. We thank our colleagues P. Boyaval, C. Fremaux, D. Romero, and E. Bech Hansen for their support and scientific contributions, and S. Moineau, V. Siksnys, and J. Banfield for their insights and expertise. This work was supported by Danisco A/S. P.H. and R.B. have submitted patent applications relating to various uses of CRISPR.

10.1126/science.1179555



CRISPR/Cas, the Immune System of Bacteria and Archaea
Philippe Horvath and Rodolphe Barrangou (January 7, 2010)
Science 327 (5962), 167-170. [doi: 10.1126/science.1179555]

Editor's Summary

CRISPR Defenses

Prokaryotes can be infected by parasites and pathogens and, like eukaryotes, have evolved systems to protect themselves. **Horvath and Barrangou** (p. 167) review a recently discovered prokaryotic "immune system" characterized by CRISPR—clustered regularly interspaced short palindromic repeats—found in most archaeal and many bacterial species. CRISPR loci harbor short sequences captured from viruses and invasive genetic elements. These sequences are transcribed, and the RNA is cleaved into short CRISPR RNAs (crRNAs) by one of a family of CRISPR-associated (cas) proteins. These crRNAs direct other cas family proteins to homologous nucleic acid targets to effect their destruction. Through its ability to impede the spread of specific nucleic acid sequences, the CRISPR/Cas systems might be exploited to block the dissemination of antibiotic-resistance markers.

This copy is for your personal, non-commercial use only.

- Article Tools** Visit the online version of this article to access the personalization and article tools:
<http://science.sciencemag.org/content/327/5962/167>
- Permissions** Obtain information about reproducing this article:
<http://www.sciencemag.org/about/permissions.dtl>

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2016 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.