CHAPTER 1

CRISPR-Cas Systems: Core Features and Common Mechanisms

Rodolphe Barrangou[1], Luciano A. Marraffini[2,3], and Erik J. Sontheimer[4]

[1]CRISPRlab, Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, NC, 27695
[2]Laboratory of Bacteriology, The Rockefeller University, New York, NY, 10065
[3]Howard Hughes Medical Institute, The Rockefeller University, New York, NY, 10065
[4]RNA Therapeutics Institute, University of Massachusetts Chan Medical School, Worcester, MA, 01605

Introduction

CRISPR loci were first described as a group of short, repetitive sequences downstream of the *iap* gene in *Escherichia coli* K-12 (1) and were later also identified in many additional bacterial and archaeal genomes (2). As opposed to other DNA repeats that are adjacent to one another, CRISPR repeats were found to be distinctively separated by similarly short sequences of unknown origin. These sequences, known as "spacers," were shown to match regions of the genomes of bacteriophages and plasmids that invade prokaryotes (3, 4, 5) (known as the target or protospacer), leading to the hypothesis that CRISPR loci could be linked to infection by these elements. In addition, a conserved set of CRISPR-associated (*cas*) genes coding for domains harbored by proteins that participate in transactions among nucleic acids (e.g., nucleases, helicases, and integrases) were often found to flank the CRISPR repeat-spacer arrays (6). Finally, the isolation and sequencing of apparent non-protein-coding RNAs from the archaeon *Archaeoglobus fulgidus* revealed that spacer sequences were transcribed and processed into small RNAs (7). All of these observations, along with in-depth bioinformatic analyses of prokaryotic genomes, allowed Koonin and Makarova to synthesize and articulate the first model for CRISPR loci as a prokaryotic defense system (8). Specifically, they proposed an RNA interference (RNAi)-based immunity, in which the small RNAs derived from the spacers would be used by Cas nucleases to find and destroy complementary transcripts, and suggested a mechanism with functional analogies to eukaryotic RNAi.

This initial phase of CRISPR studies, carried out largely *in silico*, generated intriguing and testable predictions that spawned a period of experimentation to test the bioinformatic hypotheses. In the first such study, Barrangou and colleagues showed

CRISPR: Biology and Applications, First Edition. Edited by Rodolphe Barrangou, Erik J. Sontheimer, and Luciano A. Marraffini.
 DOI: 10.1128/9781683673798.ch01

that CRISPR-Cas systems provide spacer-specific immunity against bacteriophages in *Streptococcus thermophilus* (9). More importantly, the work also produced the astonishing finding that immunity is acquired: upon infection, new spacers that match a region of the genome of the invading phage are inserted into the CRISPR array. This process, the hallmark of CRISPR immunity, is known in the field as "adaptation," as it enables the bacterial host to adapt to the environmental stress imposed by the virus. The authors also showed by gene knockout that *cas9* (originally known as *cas5* or *csn1*) is necessary for immunity and that *cas4* is required for immunization, implicating *cas* genes in CRISPR-encoded resistance. In a second study, Andersson and Banfield used metagenomic analysis of CRISPR spacer sequences to assemble viral genomes present in natural acidophilic biofilms (10). The reconstituted genomes revealed extensive recombination events that disrupted phage genome matches with the CRISPR spacers, a result showing the presence of an active arms race between CRISPR systems and phages, further supporting the role of these loci in prokaryotic defense. Later in the same year, the van der Oost group provided experimental evidence that immunity is guided by the small RNAs derived from spacers, known as CRISPR RNAs (crRNAs) (11). These researchers constructed *Escherichia coli* strains with individual deletions of the *cas* genes to identify the endoribonuclease responsible for the generation of crRNAs from a large CRISPR precursor transcript (the pre-crRNA), Cas6 (originally known as CseE). The mutant strain lacking *cas6* was unable to produce crRNAs and also failed at providing CRISPR immunity against lambda phage, demonstrating the essentiality of protospacer-complementary crRNAs for defense. The reliance of CRISPR-Cas system on this processed RNA guide also established "crRNA biogenesis" as a critical mechanistic step for immunity. Finally, Marraffini and Sontheimer showed that CRISPR loci prevent the transfer of conjugative plasmids in *Staphylococcus epidermidis* in a spacer-specific manner (12), demonstrating that CRISPR-Cas immunity extends beyond phage defense. More importantly, in part through the insertion of a self-splicing intron within the target sequence, which disrupts the DNA target but not its transcript (splicing reconstitutes the original sequence), they demonstrated that immunity can operate by targeting the DNA, as distinct from the RNA, of the invader. This finding not only contrasted with the original prediction that CRISPR-Cas systems have an RNAi-like mechanism but also prompted the proposal that the CRISPR machinery could be repurposed as an RNA-programmable, sequence-specific DNA cleavage system, perhaps even beyond prokaryotes (12). Altogether, these early mechanistic studies collectively established CRISPR-Cas as adaptive, DNA-encoded (heritable), RNA-mediated, nucleic acid-targeting immune systems (Fig. 1.1).

After these seminal studies, CRISPR research quickly accelerated. Explorations of the molecular mechanisms of immunity, as well as deep and comprehensive identification and bioinformatic analyses of CRISPR loci deposited in GenBank, led to the recognition of different CRISPR-Cas types and subtypes (13, 14, 15). Interestingly, each of the experimental studies described above investigated a different CRISPR-Cas system: the *S. thermophilus*, *E. coli*, and *S. epidermidis* experiments focused on type II-A, I-E, and III-A systems, respectively. However, the findings of these foundational studies are still relevant today because they uncovered fundamental aspects of the CRISPR-Cas immune response that are common to all types. While the chapters that follow present the details and unique features of each different CRISPR-Cas type (with the exception of type IV, about which relatively little is known), below we describe some of the unifying features of the CRISPR mechanism of action.

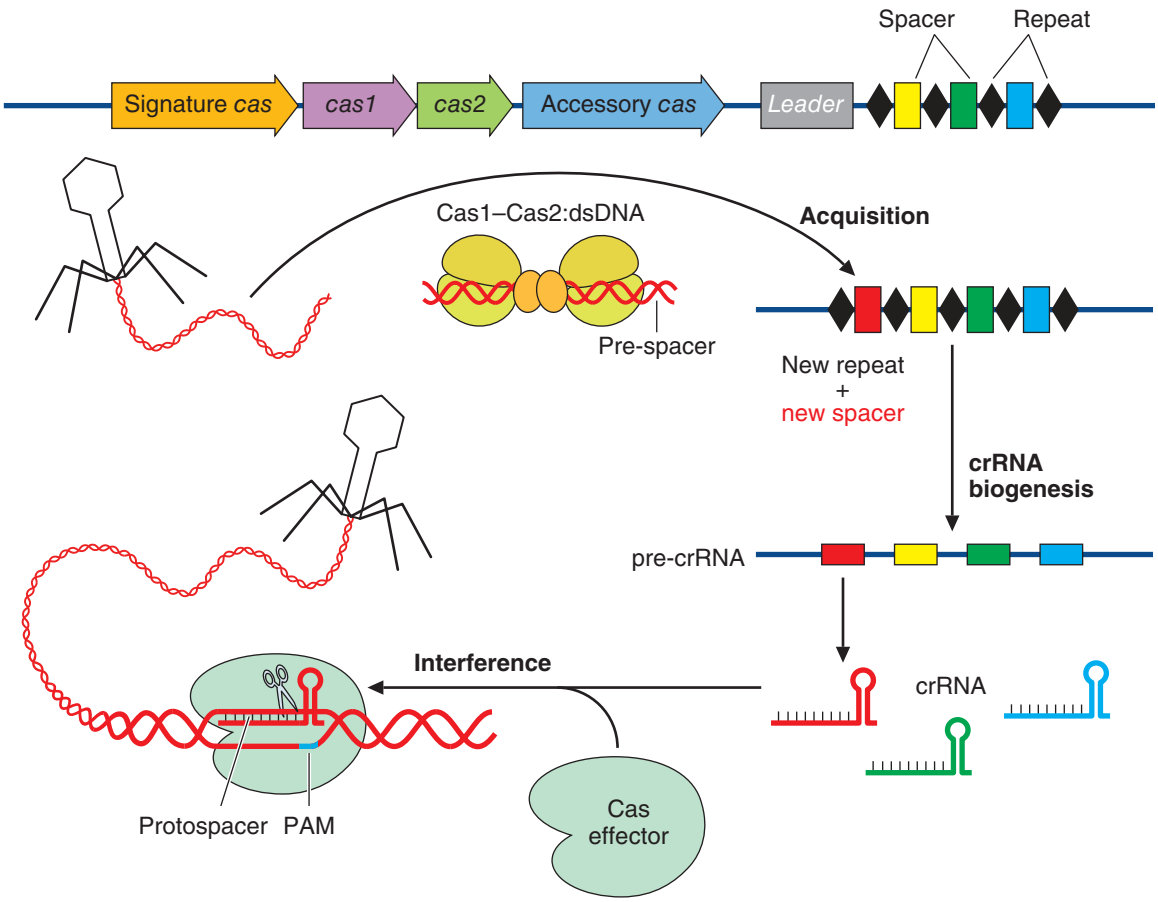

Figure 1.1 General mechanism of CRISPR-Cas immunity. The typical CRISPR locus (top) harbors the CRISPR array of repeats and spacers flanked by a set of *cas* genes. This set contains *cas1* and *cas2*, which are almost universally conserved and participate in the spacer acquisition process, as well as signature and accessory *cas* genes that determine the CRISPR type and are usually involved in the later stages of the CRISPR-Cas immune response, crRNA biogenesis and interference. During spacer acquisition, short pieces of the invader's DNA are integrated into the CRISPR array, a process that is accompanied by the duplication of a repeat sequence. During crRNA biogenesis, the CRISPR array is transcribed and processed into small RNA guides known as CRISPR RNAs (crRNAs). Finally, crRNAs guide the Cas effector complex to its complementary target nucleic acid. Here, we show the mechanism for types I, II and V CRISPR-Cas systems, where the crRNA recognizes complementary DNA target sequences that are flanked by a protospacer adjacent motif (PAM). The effector complexes of types III and VI systems use the crRNA to find complementary transcripts of the invader.

The CRISPR Array

Across all CRISPR-Cas systems, CRISPR loci are defined by their arrays of short (~40-bp) DNA repeats separated by similarly short spacer sequences. Common to the great majority of CRISPR loci is the presence of a degenerate repeat at the end of the array, with a handful of nucleotides that differ from the rest of the repeats, typically at the 3′ edge of the very last repeat (6). The repeat sequences, however, vary greatly from one locus to another, without any apparent pattern that correlates with the different CRISPR-Cas types (though with some variation that correlates with the coevolutionary dynamics between the CRISPR repeat sequences and the Cas proteins that interact with them). Indeed, it is possible to classify CRISPR loci based on their repeat sequences (16), and this can be useful for the study of "orphan" arrays (uncommon loci that lack associated *cas* genes) or arrays found in unassembled, fragmented genomes (missing their associated *cas* genes). The number of

repeats also varies substantially, with many arrays containing just a few and others a few hundred. To date, the record holder is the bacterium *Haliangium ochraceum* strain DSM 14365, which harbors 588 repeats (and 587 spacers) (17).

Spacer Acquisition

As mentioned above, one of the hallmarks of CRISPR immunity is that new spacers can be acquired upon infection from the genome of the invader (9). This creates a molecular memory of the pathogen that is used to recognize and neutralize it during subsequent attacks. New spacers are added in a polarized manner adjacent to the first repeat of the locus, also known as the "leader" end due to the presence of an A/T-rich sequence that immediately precedes this repeat. The insertion of a new spacer results in the duplication of the first repeat. This polarized form of integration creates a temporal, iterative record of infections suffered by the host, and this "recording" has been used to trace the evolutionary history of bacteria and archaea (5). Spacer acquisition thus enables an adaptive, genetically encoded, and heritable immune response that is shared across all CRISPR-Cas types.

Most CRISPR-Cas types acquire spacer sequences from DNA molecules. The exceptions are type III systems, which can convert RNA into cDNA before its integration into the CRISPR array (18), effectively capturing RNA molecules as DNA spacers (see chapter 5). In both type I and II systems, free double-stranded DNA (dsDNA) ends are the preferred substrates for the CRISPR adaptation machinery (19, 20). For both systems, the host's recombination nuclease RecBCD increases the efficiency of spacer acquisition at the free dsDNA ends. As a consequence, hot spots of acquisition are generated between these ends and the first *chi* site that stops RecBCD degradation (19, 20). Although not yet proven experimentally, it is believed that along its path of degradation from the free dsDNA end to the *chi* site, random and spontaneous disengagement of RecBCD from its substrate will presumably expose a new free dsDNA end from which spacers can be generated and acquired (21). This mode of action has important consequences for the CRISPR pathway. First, since prokaryotic chromosomes are circular but many of their invaders inject a linear DNA genome with a free dsDNA end, it provides a means to distinguish "self" from "foreign" DNA and avoid autoimmunity. Second, it allows the amplification of the CRISPR-Cas immune response through a process known as "priming" (22, 23, 24). At least in type I and II systems, the dsDNA break introduced by the Cas nucleases can lead to spacer acquisition from the free dsDNA ends that are generated (24, 25). This is particularly important to counteract the rise of escaper phages containing target mutations that prevent efficient targeting and cleavage by Cas effector nucleases. While these mutations lead to the overall failure of CRISPR immunity, in many cases this still allows a relatively low level of target cleavage, sufficient to create new free dsDNA ends and thus trigger the acquisition of new spacers to neutralize phage escapers.

At the molecular level, spacer acquisition is akin to transposon integration, although the inserted DNA is much smaller than typical transposons. The integration reaction is mediated by *cas1* and *cas2*, the only *cas* genes universally conserved across all CRISPR-Cas types (15). Cas1 and Cas2 form an integrase complex consisting of two distal Cas1 dimers bridged by a Cas2 dimer (26, 27). This complex is loaded with a prespacer DNA (the sequence that will become a spacer after integration into the CRISPR array), usually harboring 3′ overhangs, presumably originating from dsDNA ends. The mechanisms of prespacer loading are not completely elucidated and seem to be different for diverse types (27, 28). In contrast, at least in all studies to date, the integration mechanism is universal, involving the nucleophilic attack of the 3′-OH group of the prespacer overhang on the minus (bottom) strand of the first repeat, distal to the leader (Fig. 1.1). This leads to the formation of a half-site intermediate and

it is followed by a second nucleophilic attack of the other 3′-OH overhang on the plus (top) strand at the leader-repeat border. The fully integrated spacer is flanked by single-stranded repeat sequences, which are converted into double-stranded sequences, probably by a DNA polymerase gap-filling activity, to finalize the complete duplication of the repeat. Finally, there is the mechanism of selection of the first repeat for integration varies across types. In type II systems a leader-anchoring sequence promotes polarized spacer acquisition (29, 30). In contrast, type I spacer integration typically requires the bacterial integration host factor (31), which binds to the leader sequence and induces a sharp DNA bend, required for the type I Cas1-Cas2 integrase to catalyze the first integration reaction at the leader-repeat border.

CRISPR Targets

Spacer sequences can be used to find matching sequences in the large databases of DNA sequences and thus infer the targets of the CRISPR-Cas immune response. It has been estimated that only ~7% of the known spacers have homologous sequences in the databases (26,364 out of 363,460 unique spacers in reference 32), an observation that probably reflects the lack of sequence information from a majority of the prokaryotic genetic elements in general and from viruses in particular. Most of this small fraction of spacers (~96%) matched the genomes of either bacteriophages or prophages. The rest matched to plasmids (~3%), CRISPR-*cas* loci (>1%, especially *cas3*), and prokaryotic DNA that could not be identified as part of mobile genetic elements (>1%) (32). This distribution highlights the fundamental function of CRISPR-Cas systems in the defense against prokaryotic viruses (9). Similarly to phages, plasmids are a major category of prokaryotic mobile genetic elements (33), but it is not completely clear why they represent the second most abundant CRISPR target. On one hand, plasmids can carry addiction modules, such as toxin/antitoxin systems, that are costly to the host (34), which would benefit from their elimination via CRISPR immunity. On the other hand, the transfer of plasmids can be also beneficial for the recipient organism, for example, antibiotic-resistant plasmids in bacterial pathogens (35). At the population level, the limits imposed by CRISPR immunity to both phage infection and plasmid conjugation, two of the major routes of horizontal gene transfer (36), have often led to the hypothesis that CRISPR-Cas systems represent a barrier to the evolution of bacteria and archaea (12, 37). However, theoretical modeling (38) and computational estimation of the rates of horizontal gene transfer (39) have challenged this idea.

Although matched by a minority of spacers, *cas* genes represent one of the most interesting targets of CRISPR immunity. The simplest explanation for this observation is that CRISPR-Cas loci are occasionally encoded by both phage (40, 41) and plasmid (42, 43) genomes, and therefore, *cas* spacers may appear as part of the general CRISPR response against these elements. But more complex scenarios are also conceivable. For instance, some *Escherichia coli* strains carry an orphan CRISPR locus consisting of a single spacer targeting *cas* genes present in the type I-F systems of other strains (44). Importantly, the sequences of the repeats flanking the spacer are recognized and processed into a crRNA by the type I-F Cascade complex. Therefore, this orphan CRISPR can repel invasion of mobile elements that carry the type I-F *cas* locus (44), which upon entry into the host will produce a targeting complex that will turn against their own genome. This example reveals the presence of evolutionary forces that use CRISPR as a barrier against the horizontal transfer of CRISPR-*cas* loci themselves. Finally, the presence of spacers against genes that belong to core (i.e. nonmobile) prokaryotic genomic regions suggests a direct role for CRISPR against the horizontal transfer of such genes, as opposed to preventing their spread indirectly, via the attack of phages and plasmids, the main vehicles for genetic exchange (36). Supporting

this role, it has been reported that such spacers can prevent the natural transformation of both Gram-positive (45) and Gram-negative (46) bacteria. In addition, new spacers matching archaeal housekeeping genes can be acquired during interspecies mating (47), a process in which cells fuse to form a diploid state containing the full genetic repertoire of both parental cells to facilitate genetic exchange and recombination (48). Although not yet completely clear, these spacers could limit subsequent gene transfer, constraining evolution and thus facilitating archaeal speciation.

RNA-Guided Nucleases

A second hallmark of the CRISPR-Cas immune response is the employment of programmable RNA-guided nucleases. The crRNA generated from the transcription and processing of the CRISPR array is loaded into Cas proteins to form a ribonucleoprotein complex that is able to scan nucleic acids for a sequence complementary to the crRNA guide. After successful base-pairing, the target sequence is cut, with different mechanisms of cleavage for different types. For example, the crRNA is loaded into the Cas nuclease itself in types II, III, V, and VI, resulting in cleavage near or within the target sequence itself (49, 50, 51, 52, 53, 54). However, in type I systems, the crRNA is part of the Cascade complex, which lacks nuclease activity (11). This complex uses the crRNA to find the target and then recruits the helicase/nuclease Cas3, which goes on to degrade the DNA in the vicinity of the target sequence (55).

A major distinction between crRNA-guided Cas nucleases is their substrates. Types I, II, and V recognize dsDNA, a process that requires the formation of an R loop, a three-stranded nucleic acid structure composed of the target DNA:crRNA hybrid and the displaced noncomplementary DNA strand (56, 57, 58, 59, 60, 61). Recently, this form of DNA targeting has been shown (in some type I and type V instances) to specify sites of transposon insertion rather than DNA destruction (62, 63). For either of these functional outcomes, stable R-loop formation is not possible without the presence of a flanking sequence known as the protospacer adjacent motif (PAM) (53, 64, 65, 66). The PAM requirement is thought to fulfill two roles during DNA targeting. First, it facilitates the recognition of the target sequence within the invader genome. The crRNA-guided nucleases use a PAM recognition domain (60, 67, 68, 69) to scan DNA for the presence of the motif, and R-loop formation begins only when the protein-PAM interaction occurs (70, 71). Annealing of the spacer-target nucleotides immediately flanking the PAM is critical for the success of R-loop initial formation, and the presence of mismatches in this region (known as the "seed sequence") prevents targeting (64, 66). This mechanism limits the number of nonproductive attempts at R-loop formation, thus enabling rapid interrogation and identification of potential target sites within large DNA molecules. It is also suspected of providing binding energy that can be used to drive initial DNA duplex unwinding. A second function for the PAM is to avoid self-targeting of the CRISPR locus, which contains the target sequence but is flanked by a CRISPR repeat rather than a PAM. Indeed, given that the crRNA guides are fully complementary to the spacer DNA, in the absence of a PAM requirement, the spacer would be recognized as a target. However, since the spacer sequence is flanked by a repeat in the CRISPR array, which does not contain a PAM, self-targeting is not possible.

Type III and VI CRISPR-Cas systems, in contrast, use the crRNA guide to find complementary RNA transcripts produced by the invader (52, 54, 72). In these systems, target recognition triggers diverse nonspecific nuclease activities. During type III immunity, base-pairing of the crRNA with a target transcript activates the DNase activity of the Cas10

subunit (73) to destroy the template DNA. In addition, the Palm domain of Cas10 is activated and converts ATP into cyclic oligoadenylates that act as second messengers to bind and activate Csm6/Csx1 (74, 75), nonspecific RNases that induce a growth arrest of the host to assist the CRISPR defense (76, 77). Interestingly, this self-targeting is somewhat reminiscent of abortive infection, preventing infected cells from acting as viral factories. In type VI systems, target recognition by the crRNA guide within Cas13 triggers a nonspecific RNase activity of this nuclease (54, 72), also inducing a dormant state in the host that prevents the propagation of the infecting phage (78). In both types, the target RNA is cleaved shortly after its recognition, resulting in the inactivation of the nonspecific nuclease activities (54, 72, 73), presumably to prevent the irreversible destruction of the host cell. As opposed to DNA targeting systems, the crRNA-guided endoribonucleases do not appear to require a specific motif, i.e., PAM, to cleave the target RNA (79). It is believed that this may be due to the relatively easier mode of target finding (single-stranded target, no need for unwinding and R-loop formation, and high copy number of a target transcript compared to its DNA template) and also to the presence of an intrinsic limitation on self-targeting. This limitation arises because the CRISPR locus generates crRNAs but not complementary transcripts that would trigger self-immunity: "self-target" RNAs would be produced only through transcription of the CRISPR array in the opposite direction to the transcription of the crRNAs. Some level of this transcription, however, could be possible even if accidental, and both type III and VI crRNA-guided nucleases are inhibited upon recognition of self-targets. Inhibition relies on the annealing of the repeat sequence downstream of the target (the "antitag") that is complementary to the portion of the repeat sequence harbored upstream of the spacer sequence in the crRNA (the "tag") (80, 81). As a consequence of this mechanism, bona fide type III and VI targets are never flanked by antitag sequences.

Evasion of CRISPR Immunity

All bacterial and archaeal immune systems are engaged in an arms race with phages and plasmids and are therefore subject to subversion by these invaders. A great proportion of phages carry anti-CRISPR inhibitors (Acrs) (82), most of which specifically bind to and prevent the function of the crRNA-guided nucleases of each different type (covered in chapter 9). In addition to Acrs, mutations present in the phage or plasmid population that affect the recognition of the target by the Cas nucleases allow evasion of the immunity imparted by DNA-targeting CRISPR-Cas systems. As explained above, these mutations are present in either the PAM or seed sequences (64, 66). However, depending on the genetic function of the target sequence, some escape mutations are not viable, for example, if they change the structure or activity of an essential phage or plasmid gene. Therefore, some spacers mediate less "escapable," and thus more robust, immunity than others (83, 84). Importantly, spacer acquisition generates a population of bacteria that harbor numerous different spacer sequences (10, 85, 86, 87, 88). This diversity prevents the rise of escapers (89), as most phages in the population will harbor mutations that only enable escape from the immunity provided by one or two spacer sequences but eventually will infect a host cell harboring a spacer they cannot evade. Interestingly, it has been shown that phage and plasmid recombination systems can induce the accumulation of escape mutations within the targets of type I and II Cas nucleases (90, 91). Presumably, these systems can facilitate the introduction of target mutations through their involvement in the repair of the cleaved DNA. Finally, in contrast to DNA targeting CRISPR-Cas systems, the growth arrest generated during both type III and VI CRISPR immunity limits the possibilities of genetic escape (78, 79). Although there is a low number of phages

within the viral population carrying target mutations that could enable escape from recognition and targeting by Cas nucleases, these mutant phages eventually end up infecting a host in which a previous wild-type phage triggered a growth arrest. These cells are inhospitable for the propagation of any phage, including those with mutant targets. Accordingly, the diversity of the spacer repertoire seems to be less important to counteract the rise of escaper phages during type III and VI immunity (79), an observation that could have an evolutionary correlation with the very low frequencies of spacer acquisition in these systems.

Conclusions

In summary, all CRISPR-Cas systems share a common general mechanism centered on the CRISPR-*cas* locus. The CRISPR array is used to store genetic information from prokaryotic invaders, mostly phages and plasmids, through the extraction and integration of a short DNA spacer sequence from their genome. This information is passed to the Cas nucleases in the form of a crRNA guide, which is used to find and destroy specific, invasive nucleic acids. CRISPR-Cas systems are locked in a coevolutionary arms race with their targets, which have evolved a series of countermeasures to evade immunity. There are, however, many different CRISPR-Cas systems, with different features and mechanisms. The chapters that follow describe in detail the wonderful CRISPR diversity that exists in nature.

References

1. **Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A.** 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* **169:**5429–5433.
2. **Mojica FJ, Díez-Villaseñor C, Soria E, Juez G.** 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol* **36:**244–246.
3. **Bolotin A, Quinquis B, Sorokin A, Ehrlich SD.** 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading)* **151:**2551–2561.
4. **Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E.** 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60:**174–182.
5. **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 (Reading)* **151:**653–663.
6. **Jansen R, Embden JD, Gaastra W, Schouls LM.** 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43:**1565–1575.
7. **Tang T-H, Bachellerie J-P, Rozhdestvensky T, Bortolin M-L, Huber H, Drungowski M, Elge T, Brosius J, Hüttenhofer A.** 2002. Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc Natl Acad Sci U S A* **99:**7536–7541.
8. **Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV.** 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1:**7 http://dx.doi.org/10.1186/1745-6150-1-7.
9. **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.
10. **Andersson AF, Banfield JF.** 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320:**1047–1050.
11. **Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, Dickman MJ, Makarova KS, Koonin EV, van der Oost J.** 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321:**960–964.
12. **Marraffini LA, Sontheimer EJ.** 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322:**1843–1845.
13. **Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, Horvath P, Moineau S, Mojica FJM, Wolf YI, Yakunin AF, van der Oost J, Koonin EV.** 2011. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9:**467–477.
14. **Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJM, 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.
15. **Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, Charpentier E, Cheng D, Haft DH, Horvath P, Moineau S, Mojica FJM, Scott D, Shah SA, Siksnys V, Terns MP, Venclovas Č, White MF, Yakunin AF, Yan W, Zhang F, Garrett RA, Backofen R, van der Oost J, Barrangou R, Koonin EV.** 2020. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol* **18:**67–83.
16. **Nethery MA, Korvink M, Makarova KS, Wolf YI, Koonin EV, Barrangou R.** 2021. CRISPRclassify: repeat-based classification of CRISPR loci. *CRISPR J* **4:**558–574.
17. **Pourcel C, Touchon M, Villeriot N, Vernadet J-P, Couvin D, Toffano-Nioche C, Vergnaud G.** 2020. CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers. *Nucleic Acids Res* **48:**D535–D544.

18. **Silas S, Mohr G, Sidote DJ, Markham LM, Sanchez-Amat A, Bhaya D, Lambowitz AM, Fire AZ.** 2016. Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* **351**:aad4234.
19. **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.
20. **Modell JW, Jiang W, Marraffini LA.** 2017. CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* **544:**101–104.
21. **Jakhanwal S, Cress BF, Maguin P, Lobba MJ, Marraffini LA, Doudna JA.** 2021. A CRISPR-Cas9-integrase complex generates precise DNA fragments for genome integration. *Nucleic Acids Res* **49:**3546–3556.
22. **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.
23. **Swarts DC, Mosterd C, van Passel MWJ, Brouns SJJ.** 2012. CRISPR interference directs strand specific spacer acquisition. *PLoS One* **7:**e35888 http://dx.doi.org/10.1371/journal.pone.0035888.
24. **Nussenzweig PM, McGinn J, Marraffini LA.** 2019. Cas9 cleavage of viral genomes primes the acquisition of new immunological memories. *Cell Host Microbe* **26:**515–526.e6.
25. **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.
26. **Nuñez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA.** 2014. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol* **21:**528–534.
27. **Xiao Y, Ng S, Nam KH, Ke A.** 2017. How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration. *Nature* **550:**137–141.
28. **Rollie C, Graham S, Rouillon C, White MF.** 2018. Prespacer processing and specific integration in a type I-A CRISPR system. *Nucleic Acids Res* **46:**1007–1020.
29. **McGinn J, Marraffini LA.** 2016. CRISPR-Cas systems optimize their immune response by specifying the site of spacer integration. *Mol Cell* **64:**616–623.
30. **Wright AV, Doudna JA.** 2016. Protecting genome integrity during CRISPR immune adaptation. *Nat Struct Mol Biol* **23:**876–883.
31. **Nuñez JK, Bai L, Harrington LB, Hinder TL, Doudna JA.** 2016. CRISPR immunological memory requires a host factor for specificity. *Mol Cell* **62:**824–833.
32. **Shmakov SA, Sitnik V, Makarova KS, Wolf YI, Severinov KV, Koonin EV.** 2017. The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *mBio* **8**:e01397-17 http://dx.doi.org/10.1128/mBio.01397-17.
33. **Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, de la Cruz F.** 2010. Mobility of plasmids. *Microbiol Mol Biol Rev* **74:**434–452.
34. **Yang QE, Walsh TR.** 2017. Toxin-antitoxin systems and their role in disseminating and maintaining antimicrobial resistance. *FEMS Microbiol Rev* **41:**343–353.
35. **Weigel LM, Clewell DB, Gill SR, Clark NC, McDougal LK, Flannagan SE, Kolonay JF, Shetty J, Killgore GE, Tenover FC.** 2003. Genetic analysis of a high-level vancomycin-resistant isolate of *Staphylococcus aureus*. *Science* **302:**1569–1571.
36. **Thomas CM, Nielsen KM.** 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* **3:**711–721.
37. **Wheatley RM, MacLean RC.** 2021. CRISPR-Cas systems restrict horizontal gene transfer in *Pseudomonas aeruginosa*. *ISME J* **15:**1420–1433.
38. **Westra ER, Levin BR.** 2020. It is unclear how important CRISPR-Cas systems are for protecting natural populations of bacteria against infections by mobile genetic elements. *Proc Natl Acad Sci U S A* **117:**27777–27785.
39. **Gophna U, Kristensen DM, Wolf YI, Popa O, Drevet C, Koonin EV.** 2015. No evidence of inhibition of horizontal gene transfer by CRISPR-Cas on evolutionary timescales. *ISME J* **9:**2021–2027.
40. **Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD.** 2011. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res* **21:**1616–1625.
41. **Seed KD, Lazinski DW, Calderwood SB, Camilli A.** 2013. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494:**489–491.
42. **Godde JS, Bickerton A.** 2006. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J Mol Evol* **62:**718–729.
43. **Millen AM, Horvath P, Boyaval P, Romero DA.** 2012. Mobile CRISPR/Cas-mediated bacteriophage resistance in *Lactococcus lactis*. *PLoS One* **7:**e51663 http://dx.doi.org/10.1371/journal.pone.0051663.
44. **Almendros C, Guzmán NM, García-Martínez J, Mojica FJM.** 2016. Anti-*cas* spacers in orphan CRISPR4 arrays prevent uptake of active CRISPR-Cas I-F systems. *Nat Microbiol* **1:**16081 http://dx.doi.org/10.1038/nmicrobiol.2016.81.
45. **Bikard D, Hatoum-Aslan A, Mucida D, Marraffini LA.** 2012. CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* **12:**177–186.
46. **Zhang Y, Heidrich N, Ampattu BJ, Gunderson CW, Seifert HS, Schoen C, Vogel J, Sontheimer EJ.** 2013. Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Mol Cell* **50:**488–503.
47. **Turgeman-Grott I, Joseph S, Marton S, Eizenshtein K, Naor A, Soucy SM, Stachler A-E, Shalev Y, Zarkor M, Reshef L, Altman-Price N, Marchfelder A, Gophna U.** 2019. Pervasive acquisition of CRISPR memory driven by inter-species mating of archaea can limit gene transfer and influence speciation. *Nat Microbiol* **4:**177–186.
48. **Naor A, Lapierre P, Mevarech M, Papke RT, Gophna U.** 2012. Low species barriers in halophilic archaea and the formation of recombinant hybrids. *Curr Biol* **22:**1444–1448.
49. **Garneau JE, Dupuis M-È, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S.** 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468:**67–71.
50. **Gasiunas G, Barrangou R, Horvath P, Siksnys V.** 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109:**E2579–E2586.
51. **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.
52. **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.
53. **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.
54. **Abudayyeh OO, Gootenberg JS, Konermann S, Joung J, Slaymaker IM, Cox DBT, Shmakov S, Makarova KS, Semenova E, Minakhin L, Severinov K, Regev A, Lander ES, Koonin EV, Zhang F.** 2016. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**:aaf5573 http://dx.doi.org/10.1126/science.aaf5573.

55. **Westra ER, van Erp PBG, Künne T, Wong SP, Staals RHJ, Seegers CLC, Bollen S, Jore MM, Semenova E, Severinov K, de Vos WM, Dame RT, de Vries R, Brouns SJJ, van der Oost J.** 2012. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell* **46:**595–605.

56. **Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O.** 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156:**935–949.

57. **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:**147–150.

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

59. **Mulepati S, Héroux A, Bailey S.** 2014. Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* **345:**1479–1484.

60. **Yamano T, Nishimasu H, Zetsche B, Hirano H, Slaymaker IM, Li Y, Fedorova I, Nakane T, Makarova KS, Koonin EV, Ishitani R, Zhang F, Nureki O.** 2016. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* **165:**949–962.

61. **Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, Nogales E, Doudna JA.** 2016. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351:**867–871.

62. **Klompe SE, Vo PLH, Halpin-Healy TS, Sternberg SH.** 2019. Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature* **571:**219–225.

63. **Strecker J, Ladha A, Gardner Z, Schmid-Burgk JL, Makarova KS, Koonin EV, Zhang F.** 2019. RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365:**48–53.

64. **Deveau H, Barrangou R, Garneau JE, Labonté 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.

65. **Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C.** 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading)* **155:**733–740.

66. **Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJJ, 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.

67. **Sashital DG, Wiedenheft B, Doudna JA.** 2012. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol Cell* **46:**606–615.

68. **Anders C, Niewoehner O, Duerst A, Jinek M.** 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513:**569–573.

69. **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.

70. **Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA.** 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507:**62–67.

71. **Redding S, Sternberg SH, Marshall M, Gibb B, Bhat P, Guegler CK, Wiedenheft B, Doudna JA, Greene EC.** 2015. Surveillance and processing of foreign DNA by the *Escherichia coli* CRISPR-Cas system. *Cell* **163:**854–865.

72. **East-Seletsky A, O'Connell MR, Knight SC, Burstein D, Cate JHD, Tjian R, Doudna JA.** 2016. Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **538:**270–273.

73. **Kazlauskiene M, Tamulaitis G, Kostiuk G, Venclovas Č, Siksnys V.** 2016. Spatiotemporal control of type III-A CRISPR-Cas immunity: coupling DNA degradation with the target RNA recognition. *Mol Cell* **62:**295–306.

74. **Kazlauskiene M, Kostiuk G, Venclovas Č, Tamulaitis G, Siksnys V.** 2017. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* **357:**605–609.

75. **Niewoehner O, Garcia-Doval C, Rostøl JT, Berk C, Schwede F, Bigler L, Hall J, Marraffini LA, Jinek M.** 2017. Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548:**543–548.

76. **Jiang W, Samai P, Marraffini LA.** 2016. Degradation of phage transcripts by CRISPR-associated RNases enables type III CRISPR-Cas immunity. *Cell* **164:**710–721.

77. **Rostøl JT, Marraffini LA.** 2019. Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR-Cas immunity. *Nat Microbiol* **4:**656–662.

78. **Meeske AJ, Nakandakari-Higa S, Marraffini LA.** 2019. Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature* **570:**241–245.

79. **Pyenson NC, Gayvert K, Varble A, Elemento O, Marraffini LA.** 2017. Broad targeting specificity during bacterial type III CRISPR-Cas immunity constrains viral escape. *Cell Host Microbe* **22:**343–353.e3.

80. **Marraffini LA, Sontheimer EJ.** 2010. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463:**568–571.

81. **Meeske AJ, Marraffini LA.** 2018. RNA guide complementarity prevents self-targeting in type VI CRISPR systems. *Mol Cell* **71:**791–801.e3.

82. **Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR.** 2013. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493:**429–432.

83. **Paez-Espino D, Morovic W, Sun CL, Thomas BC, Ueda K, Stahl B, Barrangou R, Banfield JF.** 2013. Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat Commun* **4:**1430 http://dx.doi.org/10.1038/ncomms2440.

84. **Pyenson NC, Marraffini LA.** 2020. Co-evolution within structured bacterial communities results in multiple expansion of CRISPR loci and enhanced immunity. *eLife* **9:**e53078 http://dx.doi.org/10.7554/eLife.53078.

85. **Sun CL, Barrangou R, Thomas BC, Horvath P, Fremaux C, Banfield JF.** 2013. Phage mutations in response to CRISPR diversification in a bacterial population. *Environ Microbiol* **15:**463–470.

86. **Savitskaya E, Semenova E, Dedkov V, Metlitskaya A, Severinov K.** 2013. High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. *RNA Biol* **10:**716–725.

87. **Kieper SN, Almendros C, Behler J, McKenzie RE, Nobrega FL, Haagsma AC, Vink JNA, Hess WR, Brouns SJJ.** 2018. Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation. *Cell Rep* **22:**3377–3384.

88. **Shiimori M, Garrett SC, Graveley BR, Terns MP.** 2018. Cas4 nucleases define the PAM, length, and orientation of DNA fragments integrated at CRISPR loci. *Mol Cell* **70:**814–824.e6.

89. **van Houte S, Ekroth AKE, Broniewski JM, Chabas H, Ashby B, Bondy-Denomy J, Gandon S, Boots M, Paterson S, Buckling A, Westra ER.** 2016. The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* **532:**385–388.

90. **Roy D, Huguet KT, Grenier F, Burrus V.** 2020. IncC conjugative plasmids and SXT/R391 elements repair double-strand breaks caused by CRISPR-Cas during conjugation. *Nucleic Acids Res* **48:**8815–8827.

91. **Hossain AA, McGinn J, Meeske AJ, Modell JW, Marraffini LA.** 2021. Viral recombination systems limit CRISPR-Cas targeting through the generation of escape mutations. *Cell Host Microbe* **29:**1482–1495.e12.