Chapter 6

CRISPR: Bacteria Immune System

A. Golubov
University of Lethbridge, Lethbridge, AB, Canada

Chapter Outline

1. INTRODUCTION

All living forms from archaea and bacteria domains have various highly ingenious systems to defend themselves from foreign genetic material, which can invade them in the form of a virus particle (bacteriophage) or plasmid DNA. There is an antagonistic coevolution between bacteriophage and host cell that promotes rapid evolution of the diverse types of prokaryotic defense systems. It can be described as a combination of an arms race of defense and counter-defense between a host and an invader, and fluctuating selection on their rare genotypes [1]. To date, there are few known defensive strategies in archaea and bacteria (reviewed in Ref. [2]) that can be referred to as the multilayer prokaryotic immune system: the restriction-modification system (RMS), the adsorption inhibition, abortive infection, blocking DNA injection, and clustered regularly interspaced short palindromic repeats (CRISPRs) [3]. The initial discovery of a CRISPR structure was made accidentally in *Escherichia coli* by Ishino and colleagues in 1987 [4], but the acronym CRISPR was born in 2002 after Jansen and colleagues observed similar structures in archaeal and bacterial genomes [3]. Since 2002, our understanding of this defense system as an adaptive and heritable archaeal/bacterial immunity has made a major step forward. The highly diverse CRISPRs defense system has been found in half of the bacterial and almost all the archaeal genomes sequenced to date [5]. It is comprised of the CRISPRs and the CRISPR-associated (cas) genes (CRISPR/Cas system) that protect cells against selfish invading DNA [3].

In this chapter, I review the history of the discovery, organization in different species, and the mode of work of CRISPR, and its significance for bacterial immunity against foreign genetic material. I focus on the findings that have been made since 2000, and speculate on future perspectives of this fascinating discovery and the CRISPRs-based genome-editing technology.

2. HISTORY OF THE CRISPR/CAS DISCOVERY

The study of phage biology has a long history: bacteriophages were discovered in 1917 by the French-Canadian microbiologist Félix d'Hérelle, working at the Pasteur Institute in Paris. Since then there have been many exciting findings in that field, and it may look weird that the CRISPR/Cas immune system was not discovered before 2002. However, there are two good reasons for that: the tight regulation of the CRISPR/Cas systems in model organisms that hid it from researches, and the absence of high throughput genomic tools (such as cheap and reliable sequencing methods). Indeed,

Genome Stability. http://dx.doi.org/10.1016/B978-0-12-803309-8.00006-9

since the development of the next-generation sequencing methods, CRISPR/Cas research has been flourishing, resulting in the discovery of this highly diverse defense system in many organisms and rapid gain of knowledge about the CRISPR/Cas molecular mechanisms.

In 1987, Ishino et al. [4] were studying the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *E. coli*. Analysis of the *E. coli* chromosomal DNA fragment containing the *iap* gene revealed an unusual structure in its 3′-end flanking region: it contains "five highly homologous sequences of 29 nucleotides that were arranged as direct repeats with 32 nucleotides as spacing." At that time, due to the lack of available genome sequences, no sequence homologs have been found elsewhere in prokaryotes.

During the next 15 years, similarly organized direct repeats were found in nine archaeal genomes (eg, *Haloferax. volcanii*, *Methanocaldococcus jannaschii*, *Methanocaldococcus thermoautotrophicum*), 10 bacterial genomes (eg, *Thermotoga maritima*, *Aquifex aeolicus*, *Yersinia pestis*), and in one mitochondrial genome (*Vicia faba*) [6]. Based on the available data, Mojica et al. proposed new family of repeated short elements that are organized in clusters and named them short regularly spaced repeats (SRSRs), but their function remained unknown [6].

In 2002, Jansen et al. performed in silico analysis of a novel family of repetitive DNA sequences (SRSRs) and found that they are present in more than 40 archaeal and bacterial genomes, but absent from eukaryotes or viruses [3]. They also found that these loci were flanked on one side by a common leader sequence of 300–500 bp, and identified four genes that were invariably located adjacent to the repetitive locus, which might be an indication of their functional relationship. In order to avoid confusing nomenclature, Mojica et al. and Jansen et al. have agreed to use the acronym CRISPR, which reflects the characteristic features of this family of clustered regularly interspaced short palindromic repeats. Four adjacent genes were named cas ("CRISPR-associated"), and Jansen et al. suggested that Cas proteins play important role in the genesis of CRISPR loci. Also, they provided evidence that the putative CRISPR in *V. faba* mitochondrial genome should be considered an imperfect direct repeat and not a CRISPR.

In 2005, Bolotin et al., Mojica et al., and Pourcel et al. published their discovery that spacer sequences match viruses and plasmids, and suggested a defense function for CRISPR/Cas systems in the prevention of phage infection and plasmid conjugation [7–9].

The year of 2006 was the year of the major breakthrough: Makarova et al. incorporated all these data into a model for CRISPR/Cas immunity that is similar to the RNA interference system (RNAi) in eukaryotes and provided the first complete model for the molecular mechanism of CRISPR/Cas defense system [10]. In this model, the small CRISPR RNAs fragments had an antisense function against bacteriophage or plasmid transcripts as prokaryotic siRNAs (psiRNA), by base pairing with the target mRNAs and promoting their degradation or translation shutdown. Also, they classified numerous cas gene products into ~25 distinct protein families with several new functional and structural predictions. Cas proteins in this model have a function of the effectors of the immunity mechanism.

Two further fundamental studies uncovered two more features of CRISPR/Cas system that were not foreseen by Makarova et al.: adaptation and the targeting of the invading plasmids. In 2007, Barrangou et al. published the first evidence that the CRISPR/Cas system provides acquired resistance against viruses in prokaryotes and, thus, is adaptive [11]. To test their hypothesis about an involvement of the CRISPR/Cas system in anti-phage defense, authors infected a phage-sensitive wild-type *Streptococcus thermophilus* strain widely used in the dairy industry, DGCC7710, with two distinct but closely related virulent bacteriophages isolated from industrial yogurt samples, phage 858 and phage 2972. They have made several interesting observations about the acquired resistance: the addition of new spacers in response to phage infection seemed to be polarized toward one end of the CRISPR/Cas locus; there is similarity between the additional spacers inserted in the CRISPR/Cas locus to sequences found within the genomes of the phages used in the challenge; no particular sequence, gene, or functional group of the phages seemed to be targeted specifically. In 2008, Marraffini and Sontheimer demonstrated that the CRISPR/Cas system limits horizontal gene transfer (plasmid conjugation) in staphylococci by targeting DNA. They rigorously tested their model that consisted of wild types and deletion mutants of *Staphylococcus epidermidis* and *S. aureus* strains, and two variants of the conjugative plasmid pG0400 (wild type and mutant) [12]. The study revealed that even host cell transformation by plasmid electroporation is subject to CRISPR/Cas interference. The other important finding was that the CRISPR/Cas system provides immunity by targeting DNA, rather than RNA. Also, they were first who speculated about possibility to use the CRISPR/Cas system in genetic engineering and a clinical setting.

3. STRUCTURE OF THE CRISPR LOCI

CRISPR loci have several common features that can be found in studied organisms (Fig. 6.1): (1) multiple direct repeats with identical or nearly identical, often palindromic sequences; (2) nonrepetitive similar-sized spacer sequences; (3) a leader sequence flanking the repeats at one end; (4) the absence of functional open reading frames within the repeat arrays

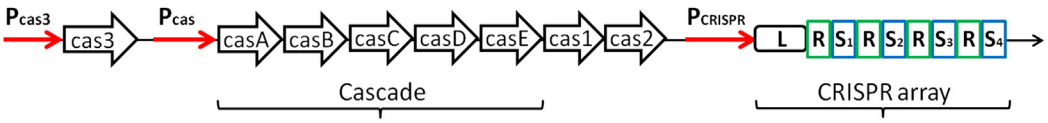

FIGURE 6.1 **cas operon and CRISPR array organization in *E. coli* K12.** The Type I system of *E. coli* K12 consists of eight cas genes and an immediately adjacent CRISPR locus. Promoters driving expression of the cas genes (P_{cas3} and P_{cas}) and the CRISPR locus (P_{CRISPR}) are indicated by *red arrows*. casABCDE genes form Cascade complex after transcription/translation. There are three distinctive elements in the CRISPR array: leader sequence (L), repeats (R), and spacers (S). One repeat and one spacer form one CRISPR unit.

and leader sequences; and (5) the genetic association of the direct repeats with cas genes. The number of CRISPR arrays varies in different organisms as well, even in different strains of the same species. It was shown that, on average, an archaeal genome contains about five CRISPR arrays, whereas three CRISPR arrays are found per bacterial genome [5]. The highest number of uninterrupted CRISPR arrays per prokaryotic genome is 20 loci in *M. jannaschii* [13].

1. Multiple direct repeats are 24–37 bp in length, they typically show weak dyad symmetry and their number per CRISPR locus can vary considerably from a few in one species to hundreds in another [14]. The largest CRISPR locus was found in *Verminephrobacter eiseniae* consisting of 245 repeats on one side and 45 repeats on the other side of an insertion sequence (IS) element [5]. Interestingly, in both archaea and bacteria, three well-separated size classes are observed: small direct repeats (24–25 bp), medium size (28–30 bp), and large (36–37 bp). The smaller direct repeats group is more represented in archaea (42% versus less than 2% for this size class in bacteria) and it is also where the differences between direct repeat and spacer size are the largest [5]. The repeat sequences have been classified in 12 different clusters, some of which are predicted to form an RNA hairpin secondary structure, whereas others are predicted to be unstructured. Kunin et al. have shown that the stems of these RNAs are well conserved and different Cas protein subtypes appeared to prefer one or more repeat types [15].
2. Nonrepetitive similar-sized spacer sequences are highly diverse elements of the CRISPR loci ranging from 26 bp to 72 bp in length. The spacers have similar lengths within a single CRISPR locus. There are no two identical spacers in the same CRISPR array, with the exception of spacer duplications in larger CRISPR arrays in some species [13]. There is a growing body of evidences suggesting that chromosomal spacers are taken up directly and, probably, randomly and nondirectionally from plasmid and bacteriophage DNA [9]. For example, the chromosomal spacers show a high level of matches (~30%) with bacteriophage or plasmid genomes in the crenarchaeal thermoacidophiles [16]. At first sight, 30% similarity to the bacteriophage or plasmid genomes looks relatively low. But the number of currently sequenced bacteriophages and plasmids is extremely small compared to the huge number of unknown mobile genetic elements that occur in nature [17].
3. A leader sequence flanking the repeats at 5′-end is an adenine/thymine (AT)-rich DNA sequence consisting of hundreds of nucleotides. It has been confirmed that it contains active promoter elements where transcription of the CRISPR array is initiated, and binding sites for regulatory proteins, such as Cas proteins involved in spacer integration, and, thus, CRISPRs adaptation [18,19].
4. So far, there is no single evidence about the presence of the functional open reading frames within the repeat arrays and leader sequences. It is not surprising, taken in mind the repetitive nature of the CRISPRs arrays and the presence of the active promoter elements in the leader sequences.
5. The possible genetic association of the direct repeats with cas genes was shown in 2002 by Jansen et al., who found that the cas genes are exclusively present in genomes that contain CRISPRs and are often located in close proximity to CRISPR arrays [3]. The cas1–4 core genes were originally marked out by Jansen et al. and are characterized by their close proximity to the CRISPR loci and their broad distribution among studied prokaryotic species. These genes are always located near a repeat locus, usually oriented head-to-tail as if cotranscribed, with the most common arrangement cas3–cas4–cas1–cas2 [3]. Haft et al. have also defined two additional core cas genes (cas5 and cas6) [14]. They have shown that the cas1–6 core gene families are not restricted to certain Cas subtypes and only two of them (cas1 and cas2) are universal. Makarova et al. have found that Cas1 (COG1518) might be the best marker of the CRISPR/Cas systems. This gene encodes a highly conserved protein and is represented in all cas neighborhoods, with the single exception of *Pyrococcus abyssii* [10]. Several studies that are based on sequence analysis predicted that Cas proteins can function as nucleases, helicases, integrases, and polymerases [20–22].

4. CRISPR/CAS CLASSIFICATION

Based on the previous observations, Haft et al. identified 45 cas gene families and eight different CRISPR/Cas subtypes that were named based on the name of organism in which the particular subtype has first been characterized: Ecoli, Ypest,

Nmeni, Dvulg, Tneap, Hmari, Apern, and Mtube [14]. There is also the RAMP module (repair-associated mysterious protein) that includes cmr1–6 genes, which is the most diverse class of cas genes, and that has been named after the RAMP superfamily.

In 2006, Makarova et al. have refined Haft's classification of 45 cas gene families by unifying them into 23 superfamilies, have tried to simplify confusing CRISPR/Cas classification, and also expanded the list of CRISPR-linked genes to include those that are found in cas operons less. In addition to the previously identified five distinct families of RAMPs, Makarova et al. detected five new ones, namely BH0337-like, y1726-like, YgcH-like, y1727-like, and MJ0978-like families. They have shown that despite the evident sequence difference, all these proteins contain the RAMP signature, the G-rich loop at the C-terminus [10].

In 2011, Makarova et al. proposed a new, polythetic classification of CRISPR/Cas systems in which the cas1 and cas2 genes form the core of three distinctive types of system (Type I, II, and III) with 10 subtypes. They have also shown that Cas1 and Cas2 are present in all CRISPR–Cas systems that are predicted to be active, and suggested to consider the Cas1/Cas2 as the information-processing subsystem that is involved in spacer integration during the adaptation stage [23]. The three types of CRISPR/Cas systems show a noticeably nonuniform distribution among the studied lineages of prokaryotes. To date, the Type II systems have been found in the bacteria only, whereas Type III systems are more common in the archaea. Also, there is a trend of overrepresentation of CRISPR in the archaea compared to the bacteria. Interestingly, the majority of sequenced archaeal genomes have more than one CRISPR/Cas system with unrelated modules within the same genome [23].

5. COMPOSITION OF THE CRISPR/CAS SYSTEMS

The *E. coli* K12 CRISPR/Cas system consists of eight cas genes: cas3, five genes designated casABCDE, cas1 (predicted integrase), and the endoribonuclease gene cas2 [21]. Five Cas proteins (Cse1, Cse2, Cas7, Cas5, and Cas6e), translated from casABCDE transcript, form a complex called Cascade. The hallmark of the Type I CRISPR/Cas systems is the cas3 gene, which encodes a protein that works as a helicase/DNase comprising of a histidine/aspartate (HD)-nuclease domain and a DExH helicase domain, in addition to genes encoding proteins that could form Cascade-like complexes [23]. These complexes contain various proteins from the RAMP superfamily (eg, Cas5, Cas6, and Cas7 families), based on the extensive sequence and structure analysis. Also, the Cascade-like complexes involved in the CRISPR/Cas/DNA interaction can contain other proteins from Cse1, Cse2, and BH0338-like families, or other, less conserved subunits [23].

Typical Type II CRISPR/Cas systems contain very large Cas9 protein, in addition to the ubiquitous Cas1 and Cas2. Cas9 has two nuclease domains and seems to be sufficient for generating crRNA and cleaving the target. It has been shown by Makarova et al. that the Type II systems include the "HNH"-type system (*Streptococcus*-like; also known as the Nmeni subtype, for *Neisseria meningitidis* serogroup A str. Z2491, or CASS4) [23]. Targeting of plasmid and phage DNA has been demonstrated in vivo using the *S. thermophilus* Type II CRISPR/Cas system, and inactivation of Cas9 has been shown to abolish interference [11,24].

The Type III CRISPR/Cas systems are the most complex and poorly understood to date, they include analogous to the Cascade complex RNA polymerase and RAMP modules, in which RAMPs seem to be involved in the processing of transcripts originated from the spacer/repeat arrays [23]. Type III systems contain, apart from the universal Cas2 protein and in addition to Cas6, at least two RAMPs that are involved in spacer/repeat transcript processing. In many studied specie, Type III CRISPR/*cas* operons lack the cas1/cas2 gene pair. Interestingly, in all these organisms, an additional either Type I or Type II CRISPR/Cas system is also present in the corresponding genome, indicating that Cas1 and Cas2 could act in trans. In all other organisms, the cas1/cas2 gene pair is combined into a single operon with the polymerase/RAMP modules, forming either a locus with the typical architecture in *S. epidermidis* and *Mycobacterium tuberculosis* (a Type IIIA module) or a distinct variant in *Halorhodospira halophila* (a Type IIIB module). The Type IIIA system in *S. epidermidis* RP62a harbors nine cas/csm genes [25]. The Type IIIB module consists of six different Cas proteins (Cmr1, Cas10, Cmr3, Cmr4, Cmr5, and Cmr6) and mature crRNA of either 39 nucleotide (nt) or 45 nt [26]. It was shown that *Sulfolobus solfataricus* contains similar crRNA-loaded Cmr complex, comprising of seven Cmr proteins (Cmr1–7), with manganese-dependent endoribonuclease activity on complementary RNA [27]. There are no other CRISPR/Cas systems found in these organisms, suggesting that the Type III system is fully functional and autonomous when the polymerase/RAMP module is present with Cas1 and Cas2 [23].

Despite the fact that the most of the CRISPR/*cas* loci can be classified into the three CRISPR/Cas types and their corresponding subtypes, there are loci that cannot fall into any specific system, even at the type level. Makarova et al. proposed to name such loci "Type U," for example, the CRISPR/Cas system in *Acidithiobacillus ferrooxidans* str. ATCC 23270 [23].

6. MOLECULAR MACHINES OF CRISPR/CAS SYSTEMS

Molecular mechanisms of the Type I CRISPR/Cas system are mediated by the multiunit Cascade complex and the Cas3 nuclease (Fig. 6.2). The Cascade complex includes Cas6e and Cas8 (also known as CasA or Cse1) subunits. The precursor crRNA (CRISPR RNAs), which is generated by transcription of the full CRISPR array, should be cut by the repeat-specific endoribonuclease Cas3 in the "crRNA biogenesis" phase. This reaction produces short crRNAs that stay associated with the Cascade complex and that are used by the complex to find a protospacer, which is a complementary sequence in the target DNA. Cas8 recognizes a short sequence motif located upstream of the target sequence recognized by the crRNA. Sequence motifs flanking the targets specified by CRISPR spacers were named as "protospacer adjacent motif," or PAM [28]. It has been shown by Semenova et al. that the PAM motif recognition is required for prevention of an autoimmune reaction for the Type I CRISPR/Cas immunity, because the absence of a PAM in the spacer/repeat array prevents the targeting of the host chromosomal DNA by their complementary crRNAs [29]. The Cascade complex binds to its target in the presence of the PAM motif, which promotes the binding and the formation of an R-loop between the crRNA spacer sequence and the target double-stranded DNA (dsDNA) [30,31]. At the final stage, the formation of the Cascade complex/target DNA structure triggers activation of the Cas3 nuclease, which introduces single-stranded DNA (ssDNA) breaks into the target DNA of bacteriophage or plasmid, thus initiating their degradation [32,33]. Semenova et al. defined a "seed" sequence within the target and showed that the first 8 bp at the 5′-end of the crRNA/DNA duplex are critical for immunity. It was shown that mutations in this region protect bacteriophages from Type I CRISPR immunity in *E. coli*. There is exclusion though: mutations in the 6th nt of the seed sequence have no effect on CRISPR immunity [29].

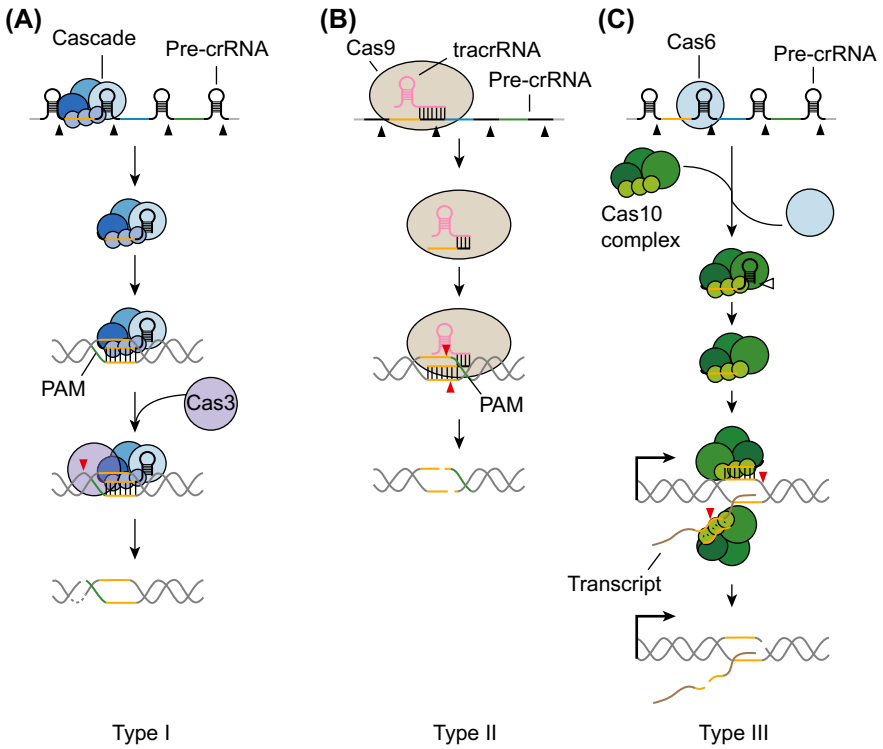

FIGURE 6.2 **Immunity mechanisms of the different CRISPR-Cas types.** (A) Type I systems. A Cas protein complex known as Cascade cleaves at the base of the stem–loop structure of each repeat in the long precursor crRNA (pre-crRNA, *black arrowheads*), which generates short crRNA guides. The Cascade– crRNA complex scans the target DNA for a matching sequence (known as protospacer), which is flanked by a protospacer-adjacent motif (PAM, in green). Annealing of the crRNA to the target strand forms an R-loop; the Cas3 nuclease is recruited and cleaves the target downstream of the PAM (*red arrowhead*) and also degrades the opposite strand. (B) Type II systems. These systems encode another small RNA known as trans-encoded crRNA (tracrRNA) which is bound by Cas9 and has regions of complementarity to the repeat sequences in the pre-crRNA. The repeat/tracrRNA/dsRNA is cleaved by RNase III to generate crRNA guides for the Cas9 nuclease (*black arrowheads*). This nuclease cleaves both strands of the protospacer/crRNA R-loop (*red arrowhead*). A PAM (in green) is located downstream of the target sequence. (C) Type III systems. Cas6 is a repeat-specific endoribonuclease that cleaves the pre-crRNA at the base of the stem–loop structure of each repeat (*black arrowhead*). The crRNA is loaded into the Cas10 complex where it is further trimmed at the 3*black arrowhead*′ end to generate a mature crRNA (*white arrowhead*). The Cas10 complex requires target transcription to cleave the nontemplate strand of the protospacer DNA and it is also capable of crRNA-guided transcript cleavage (*red arrowheads*). *Reproduced from Marraffini LA. CRISPR-Cas immunity in prokaryotes. Nature 2015;526:55–61. http://dx.doi.org/10.1038/nature15386.*

In contrast to the Type I, to perform its immunity functions, the Type II CRISPR/Cas system requires only cas9 gene, either Csn2 (Type IIA) or Cas4 (Type IIB) proteins, the presence of a targeting spacer sequence, a PAM, and two small RNAs: the crRNA and the trans-encoded crRNA (tracrRNA) [34,35]. Compared to the Type I CRISPR/Cas system, the PAM is recognized by a Cas9 PAM-binding domain and is located downstream of the target sequence [36]. The tracrRNA has two regions: one region forms a secondary structure that mediates its association with Cas9 and the other one is complementary to the repeat sequence of the CRISPR array [36]. The tracrRNA and the precursor crRNA form dsRNA, which should be digested by RNase III. Resulted cleavage products of the long CRISPR transcript are small crRNA guides [35]. Sapranauskas et al. demonstrated that there are two nuclease domains in Cas9: RuvC/RNaseH and McrA/HNH [34]. These domains are required for the Type II CRISPR/Cas immunity, which results in the introduction of crRNA-specific dsDNA breaks in the invading DNA. The tracrRNA is absolutely required for cleavage by nuclease domains of Cas9; each of these domains cleaves one DNA strand of the protospacer sequence [24,37]. It was shown that the McrA/HNH nuclease domain of Cas9 cleaves the base-pairing strand, and the RuvC/RNaseH nuclease domain cleaves the displaced strand, resulting in a blunt-end cleavage product [37]. DNA target recognition starts with the transient binding of Cas9 to PAM motifs within the target DNA, which results in the separating of the two DNA strands immediately upstream of the PAM sequences [38]. This process, in turn, triggers the creation of an R-loop and cleavage of the DNA target between 6–8 bases of the spacer sequence of the crRNA guide and the melted DNA (the "seed" region) [31].

In Type III CRISPR/Cas systems, a repeat-specific endoribonuclease, Cas6, digests the precursor crRNA at the base of a putative stem-loop structure in the pre-crRNA repeat. The result of this processing step is a sequence known as the crRNA tag, which is eight nucleotides of the repeat sequence remaining at the 5′-end of the spacer sequence in the crRNA [39]. Depending on a system subtype (IIIA or IIIB), the small crRNAs generated after Cas6 digestion are transferred to a larger complex, the Cas10/Csm or Cas10/Cmr, respectively [40]. Further, crRNAs go through a process of maturation within these complexes, by which the 3′- end is trimmed at 6-nt intervals. In the Type IIIA systems, for example, *S. epidermidis*, pre-crRNA maturation resulted in mature crRNA of two defined lengths (37 nt and 43 nt) [41]. Interestingly, The Type IIIB system of *Pyrococcus furiosus* is the first example of a prokaryotic immune system targeting RNA. It was shown that the components of the Type IIIB system can cleave RNA both in vitro and in vivo [26,42]. Recent studies of the Type IIIB system revealed that the endoribonuclease Cas6 binds crRNA at the 5′-end of the unstructured repeat sequence and cleaves it through metal-independent endoribonuclease activity, yielding 67-nt fragments. These fragments are further trimmed at the 3′-end by unknown mechanisms to 39-nt and 45-nt mature crRNA fragments that are bound to Cas proteins [22,42,43]. In the Type III CRISPR/Cas systems, the Cas10 complex contains endoribonuclease and endodeoxyribonuclease activities: subunits Csm3 and Cmr4 are responsible for cleavage of the RNA transcripts (Type IIIB system) and the palm domain of Cas10 is responsible for digestion of the nontemplate DNA strand [44–46]. To date, nobody could show that the Type III CRISPR/Cas system requires PAM for targeting activity. It was reported by Marraffini and Sontheimer that Type III system in *S. epidermidis* depends on extended base pairing between crRNA and CRISPR DNA repeats flanking the protospacer, in order to avoid autoimmunity [47]. Despite the progress in the RNA-directed immunity, the biological significance and underlying mechanisms for the RNA-targeting phenomenon has to be still elucidated.

7. CRISPR/CAS SYSTEMS AT WORK

In 2009, van der Oost et al. proposed three main stages of the development and execution of CRISPR/Cas system (Fig. 6.3): (1) adaptation (immunization), where the alien nucleic acids are encountered, and resistance is acquired by integration of an invader-derived new spacer sequence in a CRISPR array; (2) expression, where cas genes and the CRISPR array are transcribed; cas transcripts are translated, CRISPR transcripts are processed into pre-crRNAs that are subsequently digested either by a Cas6 (Type I and Type III CRISPR/Cas systems) or by an RNase III (Type II CRISPR/Cas systems); (3) interference, where the invader's nucleic acid is recognized and eliminated by the Cas/crRNA complex (crRNA recognizes complementary sequence and guides one or more Cas proteins that cleave alien DNA or RNA) [48]. We have to note that our understanding of the CRISPR/Cas systems is still partial, schemes of work are hypothetical and far from complete at the moment.

7.1 The CRISPR Adaptation

The CRISPR adaptation process was first observed in the Type II system of *S. thermophilus*, but the spacer acquisition mechanism has been studied in detail in the *E. coli* Type I CRISPR/Cas system. This is because in a report by Yosef et al., a very simple and a robust assay in *E. coli* was described that would allow to explore the process of adaptation [49]. The adaptation process can be arbitrarily divided into two stages: (1) the selection of protospacer sequences from the invader DNA and (2) integration of the protospacer sequences into the CRISPR array.

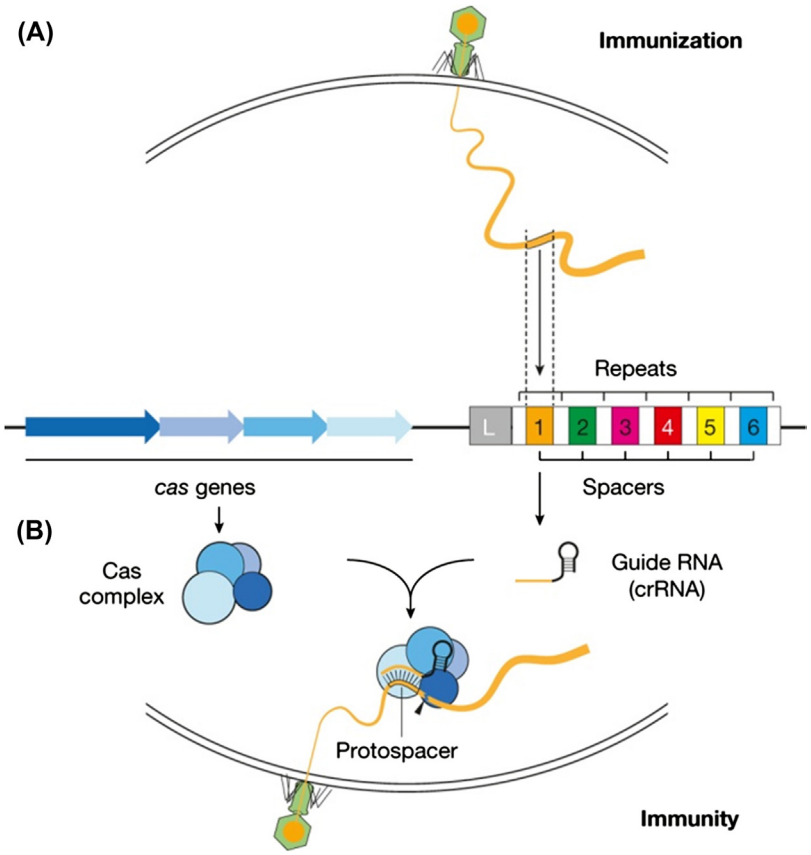

FIGURE 6.3 **Stages of CRISPR-Cas immunity.** CRISPR loci are a cluster of short DNA repeats (*white boxes*) separated by equally short spacer sequences of phage and plasmid origin (*colored, numbered boxes*). This repeat/spacer array is flanked by an operon of CRISPR-associated (cas) genes (*blue-tone arrows*) that encode the machinery for the immunization and immunity stages of the system. The CRISPR array is preceded by a leader sequence (*gray box*) containing the promoter for its expression. (A) In the immunization (adaptation) stage, spacer sequences are captured upon entry of the foreign DNA into the cell and integrated into the first position of the CRISPR array. (B) In the immunity stage (expression + interference), the spacer is used to target invading DNA that carries a cognate sequence for destruction. Spacers are transcribed and processed into small CRISPR RNAs (crRNAs) in the "crRNA biogenesis" phase. These small RNAs act as antisense guides for Cas RNA-guided nucleases (which usually form a complex) that locate and cleave the target sequence (*black arrowhead*) in the invader's genome during the "targeting" phase. *Reproduced from Marraffini LA. CRISPR-Cas immunity in prokaryotes. Nature 2015;526:55–61. http://dx.doi.org/10.1038/nature15386.*

The main aspect of the new spacer sequences selection is the ability of the prey acquisition machinery to distinguish self from nonself-DNA in order to prevent autoimmunity. Recently, it has been shown by Levy et al. in their elegant work that spacer acquisition in *E. coli* is replication-dependent, and that spacer acquisition is promoted by DNA breaks formed at stalled replication forks [50]. Numerous Chi sites (8-bp sequences) in *E. coli* restrict chromosomal hotspots of spacer acquisition, suggesting that these sites prevent spacer acquisition from self DNA. They have also shown that the autoimmunity is mediated by the RecBCD dsDNA break repair complex. They hypothesized that this model explains the strong preference of the CRISPR/Cas machinery to acquire spacers both from high-copy plasmids and from phages [50]. The second question of spacer selection relates to the PAM requirement for targeting. In *Streptococcus pyogenes*, Cas9 cleavage of bacteriophage genome requires the presence of a 5′-NGG-3′ PAM sequence immediately downstream of the viral target [35]. Heler et al. have demonstrated that Cas9 recognizes PAMs of functional spacers and, thus, provides their selection and new spacer acquisition [51]. In 2012, Datsenko et al. discovered a phenomenon that they referred to as "priming" [52]. Using phage M13 as a model system, they have shown that the presence of preexisting spacers with partial homology to the ssDNA increases the rate of spacer acquisition by several times. It has been shown that primed adaptation requires the Cascade complex, which plays dual role in immunity allowing for efficient degradation of bona fide targets and priming of mutated DNA targets [53,54].

During the second stage of the CRISPR adaptation, spacers are integrated into the CRISPR array in a reaction that is similar to retroviral integration. In 2015, it was shown by Nuñez et al. that the Cas1/Cas2 complex integrates DNA

substrates into acceptor DNA to yield products similar to those generated by retroviral integrases and transposases [55]. In this complex, Cas1 is the catalytic subunit and Cas2 substantially increases integration activity. The integration events occur preferentially at the ends of CRISPR repeats and at sequences adjacent to cruciform structures adjacent to AT-rich regions, similar to the CRISPR leader sequence [55].

7.2 The Expression Stage

The expression stage consists of three parts: regulation of the *cas* gene and CRISPR expression, Cas complex formation, and pre-crRNA maturation [48]. It has been shown by Pul et al. that transcription of pre-crRNA is directed by a promoter (P_{CRISPR}) located within the CRISPR leader sequence [18]. Transcription of the genes encoding proteins of the Cascade complex is driven by a second promoter (P_{cas}), which is located upstream of the cas genes and the activity of which resulted in a polycistronic transcript encoding Cascade (Cse1, Cse2, Cas7, Cas5, and Cas6e proteins), Cas1, and Cas2 (Fig. 6.1). Also, they have demonstrated the role of the DNA-binding protein H-NS (a global transcriptional repressor) in silencing of the CRISPR/cas promoters [18]. In contrast to other cas genes, the cas3 gene is transcribed from its own constitutive promoter (Fig. 6.1). The CRISPR/Cas system in *E. coli* K12 is also regulated by LeuO protein (a LysR-type transcription factor), which works as a transcriptional activator [56]. LeuO binds the cse1 upstream region in two sites, which flank the cse1 promoter and the H-NS nucleation site, resulting in derepression of transcription of cas genes [56]. It has been proposed by Westra et al. that regulation of transcription depends on the H–NS feature to bind AT-rich DNA, which may lead to H–NS titration from its own genome when cells encounter arrival of a mobile genetic element with AT-rich DNA [57]. This event releases repression of both the LeuO promoter (which is positively regulated by LeuO itself) and the P_{cas} promoter, initiating a CRISPR-based immune response [57].

In *E. coli* K12, the Cas6e is an endoribonuclease that cleaves the pre-crRNA in a metal-independent way, yielding mature 61-nt crRNA with an 8-nt repeat-derived 5′ handle (psi-tag) and a 21-nt 3′ handle [58]. The crRNA remains Cascade bound, forming a ribonucleoprotein complex with an attention-grabbing stoichiometry of $Cse1_1Cse2_2Cas7_6Cas5_1Cas6e_1crRNA_1$ and an asymmetrical seahorse-like shape [59]. Interestingly, in studies on different species, Cas6 homologous proteins yield similar mature crRNA with an 8-nt 5′ handle, despite the differences in protein structure and crRNA-binding ability [12,21,39,60–62].

7.3 The CRISPR Interference

CRISPR interference is a multistep process, which starts with crRNAs that forms the surveillance complex with the Cas proteins and guides it to PAM and protospacer seed regions in invading nucleic acids. The protospacer is usually dsDNA, with the exception of the Type IIIB system where the Cascade complex targets complementary ssRNA [27,63]. In the Type II and Type IIIB systems, target cleavage is carried out by the Cas/crRNA ribonucleoprotein complex, in contrast to the Type I and Type IIIA systems which require a Cas nuclease [64]. In Type I systems, the surveillance complex binds to dsDNA, recruits the Cas3 nuclease, and degrades the target [63]. In Type II systems, the Cas9/crRNA/tracrRNA complex binds to and cleaves target dsDNA [37]. The Type IIIA Csm/crRNA complex and Csm6 protein bind and most likely degrade invader dsDNA, whereas Type IIIB Cmr/crRNA complex cleaves complementary RNA [65]. The Csm complex in the Type IIIA system includes six different proteins but the nuclease is not yet identified [66]. The Cmr complex in the Type IIIB system includes six or seven different proteins and the Cmr4 protein works as a nuclease [67]. It has been shown that in *Thermus thermophilus* and *S. thermophiles*, the Csm complex targets RNA, and in *T. thermophilus*, which harbors both Type IIIA and IIIB systems, the Csm and Cmr complexes share crRNA [45,46]. In *S. islandicus*, Cmr complex targets both RNA and DNA [68,69]. Another interesting CRISPR-Cas interference mechanism was recently found in *S. epidermidis*: it can prevent lytic infection but tolerate lysogenization by temperate phages. It was shown that conditional tolerance is achieved through transcription-dependent DNA targeting, and ensures that targeting is resumed upon induction of the prophage lytic cycle [70].

8. OTHER ROLES OF THE CRISPR/CAS SYSTEMS

There is a growing body of evidence suggesting that the CRISPR/Cas system, besides its immunity function, can be a part of other cellular processes such as DNA repair and regulation of virulence.

Taking in mind the ability of the CRISPR/Cas systems to shape a bacterial genome landscape by acquisition of new spacers, it was quite obvious to hypothesize that these systems might have an impact on the stability and evolution of bacterial genomes. Indeed, recent study of *S. thermophilus* revealed that the CRISPR/Cas systems target mobile genetic elements

(bacteriophages, transposons, and plasmids), which likely contributed to gene acquisition and loss during evolutionary adaptation to milk, thus limiting genetic diversity and stabilizing of the *S. thermophilus* genome [71]. On the contrary, the genome analysis of *T. maritima* MSB8 and *Thermotoga neapolitana* NS-E provided evidence that the CRISPR/Cas systems might be a cause of numerous CRISPR-associated large-scale DNA rearrangements that destabilize and reshape genomes [72].

It was shown that the Cas1 protein of *E. coli* interacts with RecB, RecC, and RuvB, it can process single-stranded and branched DNA species, replication forks and 5′ flaps [73]. In *Francisella novicida*, Cas9 protein uses a unique, small, CRISPR/Cas-associated RNA (scaRNA) to repress transcription of a bacterial lipoprotein (FTN_1103). As bacterial lipoproteins trigger a proinflammatory innate immune response in a host, CRISPR/Cas-mediated transcriptional repression of FTN_1103 is important for *F. novicida* to reduce this host response and promote virulence [74]. It has been demonstrated in *Campylobacter jejuni* that inactivation of the Type II CRISPR/Cas marker gene csn1 effectively reduced virulence in primarily cst-II-positive *C. jejuni* isolates [75]. cas2 mutants in *Legionella pneumophila*, although they grew typically in macrophages, were significantly impaired for infection of both *Hartmannella* and *Acanthamoeba* species. Given that infection of amebae is critical for *L. pneumophila* persistence in water systems, these data indicate that cas2 might play a role in the transmission of Legionnaires' disease [76].

To date, there is not enough data to draw a conclusion on whether the CRISPR/Cas systems are mainly involved in the bacterial immunity. Above-mentioned examples raise interesting questions about the evolution of CRISPR/Cas function, which require more in-depth research.

9. CONCLUSION

CRISPR/Cas research experienced tremendous boost during the last decade. It should be appreciated that the CRISPR/Cas discovery has had a huge impact on bacteriology and genetic engineering, which can be compared to, for example, discovery of the polymerase chain reaction or development of the next-generation sequencing technology. However, many molecular details and mechanisms of action CRISPR/Cas systems remain to be determined. It would be very interesting to see whether cross-talk exists in species with multiple CRISPR/Cas systems, their regulation, effect on virulence in the case of animal pathogens, and many others. Another baffling problem of CRISPR/Cas systems is their high diversity: there are many subtypes of the "main" three CRISPR types with different sets of Cas proteins, structural organization, and regulation. The biological significance of this diversity remains to be determined.

GLOSSARY

Dyad symmetry Dyad symmetry refers to two areas of a DNA strand whose base-pair sequences are inverted repeats of each other. They are often described as palindromes. For example, the following shows dyad symmetry between sequences GAATAC and GTATTC which are reverse complements of each other.

Innate immune response Innate immune response is the first line of defense against invading microbial pathogens and relies on a large family of pattern recognition receptors (PRRs), which detect distinct evolutionarily conserved structures on pathogens, termed pathogen-associated molecular patterns (PAMPs). Among the PRRs, the Toll-like receptors have been studied most extensively.

IS element IS element (also known as an IS, an insertion sequence element, or an insertion sequence) is a short DNA sequence that acts as a simple transposable element. IS have two major characteristics: they are small relative to other transposable elements (generally about 700–2500 bp in length) and only code for proteins implicated in the transposition activity (they are thus different from other transposons, which also carry accessory genes such as antibiotic-resistance genes).

Next-generation sequencing (NGS) Next-generation sequencing also known as high-throughput sequencing, is the catch-all term used to describe a number of different modern sequencing technologies, including Illumina (Solexa) sequencing, Roche 454 sequencing, Ion torrent (Proton/PGM) sequencing, and SOLiD sequencing. These technologies allow us to sequence DNA and RNA much more quickly and cheaply than the previously used Sanger sequencing, and as such have revolutionized the study of genomics and molecular biology.

Polythetic classification Polythetic classification is defined in terms of a broad set of criteria that are neither necessary nor sufficient. Each member of the category must possess a certain minimal number of defining characteristics, but none of the features has to be found in each member of the category. This way of defining classes is associated with Wittgenstein's concept of "family resemblances."

Protospacer Protospacer is the sequence complementary to the crRNA (CRISPR RNA) spacer.

RNA interference (RNAi) RNA interference is an important pathway that is used in many different organisms to regulate gene expression. This is a biological process in which small RNA molecules inhibit gene expression, typically by causing the destruction of specific mRNA molecules. Historically, it was known by other names, including *cosuppression*, *posttranscriptional gene silencing* (PTGS), and *quelling*. Two types of small RNA molecules—microRNA (miRNA) and small interfering RNA (siRNA)—are central to RNA interference.

Seed region Seed region is a noncontiguous 7-nt region of a protospacer (positions 1–5, 7, and 8 of the spacer) near the 5′-end of the crRNA.

Virulence Virulence is the degree of pathogenicity within a group or species of pathogens as indicated by case fatality rates and/or the ability of the organism to invade the tissues of the host.

LIST OF ACRONYMS AND ABBREVIATIONS

bp Base pairs
cas genes CRISPR-associated genes
CRISPRs Clustered regularly interspaced short palindromic repeats
crRNA CRISPR RNA
dsDNA Double-stranded DNA
dsRNA Double-stranded RNA
nt Nucleotides
PAM Protospacer adjacent motif
psiRNA Prokaryotic small interfering RNA
RAMP Repair-associated mysterious protein
RMS Restriction-modification system
RNAi RNA interference
SRSRs Short regularly spaced repeats
ssDNA Single-stranded DNA
tracrRNA Transencoded crRNA

REFERENCES

[1] Hall AR, Scanlan PD, Morgan AD, Buckling A. Host-parasite coevolutionary arms races give way to fluctuating selection. Ecol Lett July 2011;14(7):635–42.
[2] Labrie SJ, Samson JE, Moineau S. Bacteriophage resistance mechanisms. Nat Rev Microbiol May 2010;8(5):317–27.
[3] Jansen R, Embden JD, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol March 2002;43(6):1565–75.
[4] Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J Bacteriol December 1987;169(12):5429–33.
[5] Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinf 2007;8:172.
[6] Mojica FJ, Diez-Villasenor C, Soria E, Juez G. Biological significance of a family of regularly spaced repeats in the genomes of archaea, bacteria and mitochondria. Mol Microbiol April 2000;36(1):244–6.
[7] Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology August 2005;151(Pt 8):2551–61.
[8] Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol February 2005;60(2):174–82.
[9] 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 March 2005;151(Pt 3):653–63.
[10] Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. 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 2006;1:7.
[11] Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science March 23, 2007;315(5819):1709–12.
[12] Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science December 19, 2008;322(5909):1843–5.
[13] Lillestol RK, Redder P, Garrett RA, Brugger K. A putative viral defence mechanism in archaeal cells. Archaea August 2006;2(1):59–72.
[14] Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol November 2005;1(6):e60.
[15] Kunin V, Sorek R, Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol 2007;8(4):R61.
[16] 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 February 2009;37(Pt 1):23–8.
[17] Snyder JC, Bateson MM, Lavin M, Young MJ. Use of cellular CRISPR (clusters of regularly interspaced short palindromic repeats) spacer-based microarrays for detection of viruses in environmental samples. Appl Environ Microbiol November 2010;76(21):7251–8.
[18] 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 March 2010;75(6):1495–512.
[19] Al-Attar S, Westra ER, van der Oost J, Brouns SJ. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. Biol Chem April 2011;392(4):277–89.
[20] 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 June 10, 2009;17(6):904–12.
[21] Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science August 15, 2008;321(5891):960–4.

[22] Carte J, Pfister NT, Compton MM, Terns RM, Terns MP. Binding and cleavage of CRISPR RNA by Cas6. RNA November 2010;16(11):2181–8.
[23] Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, et al. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol June 2011;9(6):467–77.
[24] 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 November 4, 2010;468(7320):67–71.
[25] Hatoum-Aslan A, Maniv I, Samai P, Marraffini LA. Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. J Bacteriol January 2014;196(2):310–7.
[26] 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 November 25, 2009;139(5):945–56.
[27] Zhang J, Rouillon C, Kerou M, Reeks J, Brugger K, Graham S, et al. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. Mol Cell February 10, 2012;45(3):303–13.
[28] Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology March 2009;155(Pt 3):733–40.
[29] 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 June 21, 2011;108(25):10098–103.
[30] Rutkauskas M, Sinkunas T, Songailiene I, Tikhomirova MS, Siksnys V, Seidel R. Directional R-loop formation by the CRISPR-Cas surveillance complex Cascade provides efficient off-target site rejection. Cell Rep March 3, 2015. http://dx.doi.org/10.1016/j.celrep.2015.01.067.
[31] Szczelkun MD, Tikhomirova MS, Sinkunas T, Gasiunas G, Karvelis T, Pschera P, et al. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc Natl Acad Sci USA July 8, 2014;111(27):9798–803.
[32] Hochstrasser ML, Taylor DW, Bhat P, Guegler CK, Sternberg SH, Nogales E, et al. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proc Natl Acad Sci USA May 6, 2014;111(18):6618–23.
[33] Mulepati S, Bailey S. In vitro reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. J Biol Chem August 2, 2013;288(31):22184–92.
[34] 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 November 2011;39(21):9275–82.
[35] Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature March 31, 2011;471(7340):602–7.
[36] Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature September 25, 2014;513(7519):569–73.
[37] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science August 17, 2012;337(6096):816–21.
[38] Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature March 6, 2014;507(7490):62–7.
[39] Carte J, Wang R, Li H, Terns RM, Terns MP. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev December 15, 2008;22(24):3489–96.
[40] Sokolowski RD, Graham S, White MF. Cas6 specificity and CRISPR RNA loading in a complex CRISPR-Cas system. Nucleic Acids Res June 2014;42(10):6532–41.
[41] Hatoum-Aslan A, Samai P, Maniv I, Jiang W, Marraffini LA. A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. J Biol Chem September 27, 2013;288(39):27888–97.
[42] 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 February 10, 2012;45(3):292–302.
[43] Wang R, Preamplume G, Terns MP, Terns RM, Li H. Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. Structure February 9, 2011;19(2):257–64.
[44] Samai P, Pyenson N, Jiang W, Goldberg GW, Hatoum-Aslan A, Marraffini LA. Co-transcriptional DNA and RNA Cleavage during type III CRISPR-Cas immunity. Cell May 21, 2015;161(5):1164–74.
[45] Staals RH, Zhu Y, Taylor DW, Kornfeld JE, Sharma K, Barendregt A, et al. RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. Mol Cell November 20, 2014;56(4):518–30.
[46] Tamulaitis G, Kazlauskiene M, Manakova E, Venclovas C, Nwokeoji AO, Dickman MJ, et al. Programmable RNA shredding by the type III-A CRISPR-Cas system of *Streptococcus thermophilus*. Mol Cell November 20, 2014;56(4):506–17.
[47] Marraffini LA, Sontheimer EJ. Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature January 28, 2010;463(7280):568–71.
[48] van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ. CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem Sci August 2009;34(8):401–7.
[49] Yosef I, Goren MG, Qimron U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. Nucleic Acids Res July 2012;40(12):5569–76.
[50] Levy A, Goren MG, Yosef I, Auster O, Manor M, Amitai G, et al. CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature April 23, 2015;520(7548):505–10.
[51] Heler R, Samai P, Modell JW, Weiner C, Goldberg GW, Bikard D, et al. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature March 12, 2015;519(7542):199–202.
[52] 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.

[53] Blosser TR, Loeff L, Westra ER, Vlot M, Kunne T, Sobota M, et al. Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. Mol Cell April 2, 2015;58(1):60–70.
[54] Fineran PC, Gerritzen MJ, Suarez-Diez M, Kunne T, Boekhorst J, van Hijum SA, et al. Degenerate target sites mediate rapid primed CRISPR adaptation. Proc Natl Acad Sci USA April 22, 2014;111(16):E1629–38.
[55] Nunez JK, Lee AS, Engelman A, Doudna JA. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. Nature March 12, 2015;519(7542):193–8.
[56] 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 September 2010;77(6):1380–93.
[57] 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.
[58] Jore MM, Brouns SJ, van der Oost J. RNA in defense: CRISPRs protect prokaryotes against mobile genetic elements. Cold Spring Harb Perspect Biol June 2012;4(6).
[59] Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol May 2011;18(5):529–36.
[60] Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science September 10, 2010;329(5997):1355–8.
[61] 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 June 17, 2011;286(24):21643–56.
[62] Terns RM, Terns MP. The RNA- and DNA-targeting CRISPR-Cas immune systems of *Pyrococcus furiosus*. Biochem Soc Trans December 2013;41(6):1416–21.
[63] Westra ER, van Erp PB, Kunne 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 June 8, 2012;46(5):595–605.
[64] Reeks J, Naismith JH, White MF. CRISPR interference: a structural perspective. Biochem J July 15, 2013;453(2):155–66.
[65] Westra ER, Buckling A, Fineran PC. CRISPR-Cas systems: beyond adaptive immunity. Nat Rev Microbiol May 2014;12(5):317–26.
[66] Rouillon C, Zhou M, Zhang J, Politis A, Beilsten-Edmands V, Cannone G, et al. Structure of the CRISPR interference complex CSM reveals key similarities with cascade. Mol Cell October 10, 2013;52(1):124–34.
[67] Benda C, Ebert J, Scheltema RA, Schiller HB, Baumgartner M, Bonneau F, et al. Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4. Mol Cell October 2, 2014;56(1):43–54.
[68] 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 March 2013;87(5):1088–99.
[69] Peng W, Feng M, Feng X, Liang YX, She Q. An archaeal CRISPR type III-B system exhibiting distinctive RNA targeting features and mediating dual RNA and DNA interference. Nucleic Acids Res January 2015;43(1):406–17.
[70] Goldberg GW, Jiang W, Bikard D, Marraffini LA. Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. Nature October 30, 2014;514(7524):633–7.
[71] Selle K, Klaenhammer TR, Barrangou R. CRISPR-based screening of genomic island excision events in bacteria. Proc Natl Acad Sci USA June 30, 2015;112(26):8076–81.
[72] DeBoy RT, Mongodin EF, Emerson JB, Nelson KE. Chromosome evolution in the Thermotogales: large-scale inversions and strain diversification of CRISPR sequences. J Bacteriol April 2006;188(7):2364–74.
[73] 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 January 2011;79(2):484–502.
[74] Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature May 9, 2013;497(7448):254–7.
[75] Louwen R, Horst-Kreft D, de Boer AG, van der Graaf L, de Knegt G, Hamersma M, et al. A novel link between *Campylobacter jejuni* bacteriophage defence, virulence and Guillain-Barre syndrome. Eur J Clin Microbiol Infect Dis February 2013;32(2):207–26.
[76] Gunderson FF, Cianciotto NP. The CRISPR-associated gene cas2 of *Legionella pneumophila* is required for intracellular infection of amoebae. MBio 2013;4(2):e00074–13.
[77] Marraffini LA. CRISPR-Cas immunity in prokaryotes. Nature October 1, 2015;526(7571):55–61.