CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea

Rotem Sorek,[1] C. Martin Lawrence,[2,3] and Blake Wiedenheft[4]

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel; email: rotem.sorek@weizmann.ac.il

[2]Thermal Biology Institute, [3]Department of Chemistry and Biochemistry, and [4]Department of Immunology and Infectious Diseases, Montana State University, Bozeman, Montana 59717; email: lawrence@chemistry.montana.edu, blake.wiedenheft@montana.edu

Annu. Rev. Biochem. 2013. 82:237–66

First published online as a Review in Advance on March 11, 2013

The *Annual Review of Biochemistry* is online at biochem.annualreviews.org

This article's doi: 10.1146/annurev-biochem-072911-172315

Keywords

bacterial immunity, bacteriophage, crRNA, RNA interference, small RNAs

Abstract

Effective clearance of an infection requires that the immune system rapidly detects and neutralizes invading parasites while strictly avoiding self-antigens that would result in autoimmunity. The cellular machinery and complex signaling pathways that coordinate an effective immune response have generally been considered properties of the eukaryotic immune system. However, a surprisingly sophisticated adaptive immune system that relies on small RNAs for sequence-specific targeting of foreign nucleic acids was recently discovered in bacteria and archaea. Molecular vaccination in prokaryotes is achieved by integrating short fragments of foreign nucleic acids into a repetitive locus in the host chromosome known as a CRISPR (clustered regularly interspaced short palindromic repeat). Here we review the mechanisms of CRISPR-mediated immunity and discuss the ecological and evolutionary implications of these adaptive defense systems.

Contents

INTRODUCTION

In the late 1800s, Ernest Hanbury Hankin (1) reported that water from the Ganges and Yamuna rivers in India contained an antibacterial agent that killed *Vibrio cholerae*. These filterable agents, later termed bacteriophages (from "bacteria" and the Greek word *phagein*, "to devour"), were heralded as a potential treatment for diseases. Although phages have yet to reach their therapeutic potential in clinical settings, the importance of bacteriophages in environmental and medical science is currently reaching a new crescendo. In the 1980s, marine virologists reported that one liter of sea water contains approximately ten billion bacteriophages, and today these viruses are generally considered the most abundant and diverse biological entities on Earth (2–4). The selective pressures imposed by these viral predators have a profound impact on the composition and the behavior of microbial communities in every ecological setting (5), and microbial hosts have evolved various mechanisms to evade infection (6–9). Historically our appreciation for microbial immune systems had been restricted to innate defense mechanisms (e.g., restriction modification and receptor switching), but a nucleic acid–based adaptive immune system was recently discovered (10–13). Bacteria and archaea acquire resistance to viral and plasmid challengers by integrating short fragments of foreign nucleic acid into the host chromosome at one end of a repetitive element known as a CRISPR (clustered regularly interspaced short palindromic repeat). CRISPR-mediated adaptive immunity proceeds in three distinct stages: acquisition of foreign DNA, CRISPR RNA (crRNA) biogenesis, and target interference (**Figure 1*a***). Although these three basic stages appear to be common to all CRISPR systems, CRISPR loci and the proteins that mediate each stage of adaptive immunity are remarkably diverse (**Figure 1*b***). Here we review the functional diversity among different versions of this immune system and discuss the evolutionary implications of this rapidly evolving, heritable immune system on microbial evolution.

CRISPR DESIGN AND DISTRIBUTION

CRISPRs are a diverse family of DNA repeats that all share a common architecture. Each CRISPR locus consists of a series of short repeat sequences [typically 20–50 base pairs (bp) long] separated by unique spacer sequences of a similar length (**Figure 1*a***). The repeat sequences within a CRISPR locus are

conserved, but repeats in different CRISPR loci can vary in both sequence and length (14). Phylogenetic analyses of CRISPR repeat sequences have shown that CRISPRs can be organized into clusters based on the sequence similarity of their repeat sequences. Some repeats are palindromic and are predicted to generate RNAs with stable hairpin structures, whereas others are predicted to be unstructured (**Figure 1*b***) (14). Despite the extreme diversity of CRISPR repeat sequences, most repeats have a conserved GAAA(C/G) motif at the 3′ end, which may serve as a binding site for one or more of the conserved Cas proteins (14–16).

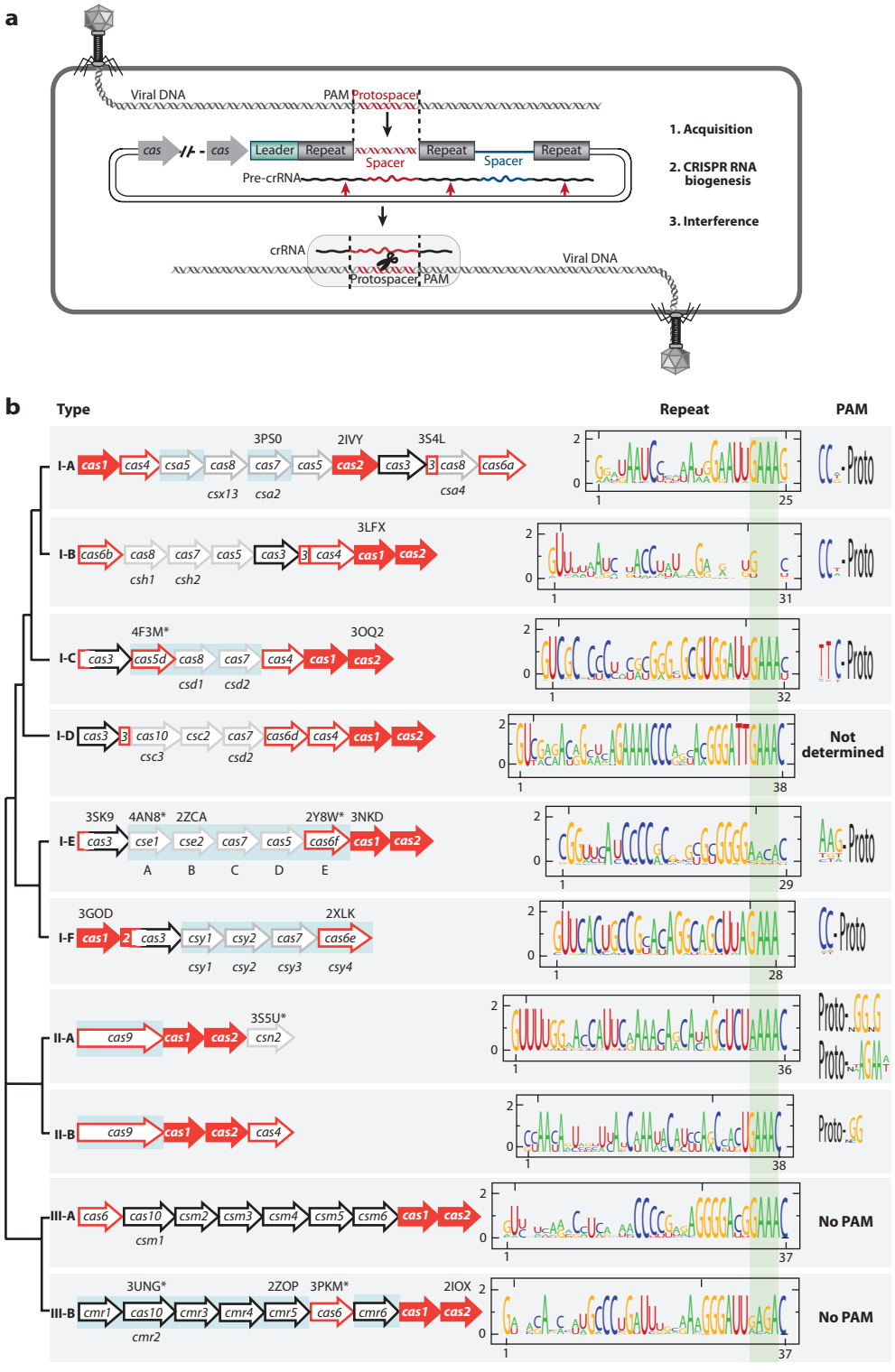
In addition to repeat and spacer sequence diversity, the number of CRISPR loci and the length of each locus are also variable. It is not uncommon for a single prokaryotic chromosome to contain multiple CRISPR loci (e.g., 18 CRISPR loci in *Methanocaldococcus* sp. FS406-22), and some of these loci can be thousands of base pairs in length (hundreds of repeat-spacer units). The number of distinct CRISPR loci and the length of these repetitive arrays do not correlate with genome size; some of the smallest microbial genomes (e.g., *Nanoarchaeum equitans*) contain multiple CRISPR loci, and CRISPRs in some genomes account for more than 1% of the chromosome (e.g., *Sulfolobus solfataricus*).

The repeat-spacer-repeat pattern now considered to be the defining characteristic of a CRISPR locus was initially described in *Escherichia coli* in 1987 (17). However, the prevalence and phylogenetic distribution of these repetitive elements were not appreciated for more than a decade (16). Computational methods for detecting these repeat patterns have been developed, and there are currently two web-based utilities (CRISPRdb and CRISPI) dedicated to the identification and annotation of CRISPRs and CRISPR-associated (*cas*) genes (18, 19). Interestingly, CRISPRs are unevenly distributed between bacteria and archaea. Currently, CRISPR loci have been identified in ~90% of the archaeal genomes and ~50% of the bacterial genomes. Although the biological basis for this skewed distribution remains speculative, an assessment of 24 *Enterococcus faecalis* genome sequences revealed an inverse correlation between the presence of a CRISPR/Cas locus and antibiotic resistance (20).

An adenine and thymine (AT)-rich sequence called a leader often flanks CRISPR loci. Comparative analyses have shown that spacer sequences nearest the leader are most diverse, whereas repeats farthest from the leader (in the region known as the trailer) are often degenerate (21, 22). Although the function of CRISPRs was unknown at the time of this initial observation, the leader-end diversity and trailer-end degeneracy indicated these loci had polarity defined by the position of the leader. We now know that the leader sequences contain promoter elements (23–26) and binding sites for regulatory proteins (25, 26) critical to crRNA expression and new sequence acquisition (27).

CRISPRs AND THEIR ASSOCIATED PROTEIN MACHINERY

In addition to the leader sequence, comparative analyses have also identified a variable cassette of *cas* genes, which is typically located adjacent to a CRISPR locus (**Figure 1*b***). Four *cas* genes were initially identified in genomes containing CRISPRs (21), but accumulating genome sequences and the implementation of increasingly sophisticated search methods have led to the identification of ~45 different gene families commonly found in association with CRISPRs (28). Six of these *cas* genes (*cas1–cas6*) are widely conserved and are considered core *cas* genes, but only *cas1* and *cas2* are universally conserved in genomes that contain CRISPR loci (28, 29). *cas1* is a hallmark of this immune system, and phylogenetic analysis of *cas1* sequences suggests several distinct versions of CRISPR systems exist (28, 29). Each of these different phylotypes is defined by a unique composition and conserved arrangement of *cas* genes. Remarkably, this *cas* gene–based classification appears to correlate well with a CRISPR repeat–based classification, suggesting that the Cas proteins interact with specific sets of CRISPR loci (14, 29).


a
Viral DNA
PAM
Protospacer
cas
cas
Leader
Repeat
Spacer
Repeat
Spacer
Repeat
Pre-crRNA
1. Acquisition
2. CRISPR RNA biogenesis
3. Interference
crRNA
Protospacer
PAM
Viral DNA
b
Type
Repeat
PAM
I-A
cas1
cas4
csa5
cas8
csx13
cas7
csa2
cas5
cas2
cas3
cas8
csa4
cas6a
3PS0
2IVY
3S4L
I-B
cas6b
cas8
csh1
cas7
csh2
cas5
cas3
cas4
cas1
cas2
3LFX
I-C
cas3
cas5d
csd1
cas8
csd2
cas7
cas4
cas1
cas2
4F3M*
3OQ2
I-D
cas3
cas10
csc3
csc2
cas7
csd2
cas6d
cas4
cas1
cas2
Not determined
I-E
cas3
cse1
A
cse2
B
cas7
C
cas5
D
cas6f
E
cas1
cas2
3SK9
4AN8*
2ZCA
2Y8W*
3NKD
I-F
cas1
cas3
csy1
csy1
csy2
csy2
cas7
csy3
cas6e
csy4
3GOD
2XLK
II-A
cas9
cas1
cas2
csn2
3S5U*
II-B
cas9
cas1
cas2
cas4
III-A
cas6
cas10
csm1
csm2
csm3
csm4
csm5
csm6
cas1
cas2
No PAM
III-B
cmr1
cas10
cmr2
cmr3
cmr4
cmr5
cas6
cmr6
cas1
cas2
3UNG*
2ZOP
3PKM*
2IOX
No PAM

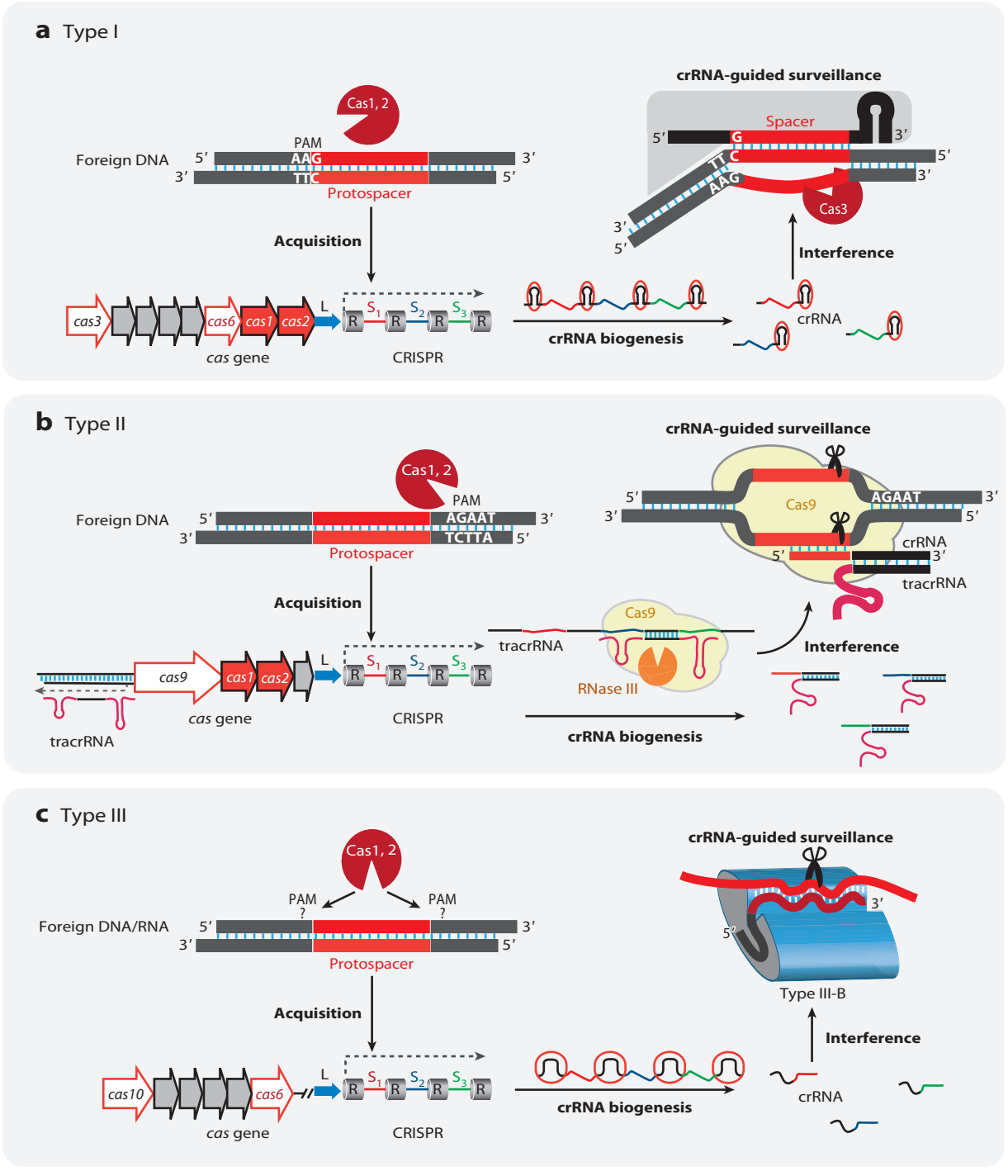
The different CRISPR repeat clusters were initially numbered 1 to 12 (14), whereas the different *cas* systems were originally named after a representative organism, using a three letter code (28). For example, the Cas system in *E. coli* K12 was designated *cse* (i.e., CRISPR system *E. coli*), and each *cse* gene was assigned a number according to its position in the *cas* gene cluster (e.g., cse1, cse2). The *cas* genes in the other systems were named using a similar strategy. However, some of the *cas* gene families were later determined to be orthologous and renamed using a "clusters of orthologous groups" classification scheme (29). These pioneering phylogenetic studies were critical to establishing a foundation for biochemical and mechanistic investigation, but the diversity of *cas* genes and their association with different CRISPR repeat clusters have made it challenging to arrive at a common vernacular that is easy to interpret. A newly proposed classification scheme integrates *cas* gene and CRISPR repeat phylogenies (31). With this approach, three major types of CRISPR/Cas systems have been delineated, and each of these major types can be divided into subtypes (i.e., type IA–F, type IIA–B, and type IIIA–B).

Type I CRISPR-Associated Systems

Type I CRISPR/Cas systems are widely distributed in bacteria and archaea (31). Type I systems encompass six distinct subtypes (A–F), all of which encode a *cas3* gene (**Figures 1*b*** and **2*a***). Cas3 contains an N-terminal HD phosphohydrolase domain and a C-terminal DExH helicase domain (29, 31). In some type I systems (subtypes A, B, and D), separate genes encode the nuclease and helicase domains, but in all of these systems these two domains are anticipated to work together by cleaving (HD domain) and unwinding (helicase domain) dsDNA (double-stranded DNA) targets for processive degradation. However, Cas3 cannot identify foreign DNA, and by itself it cannot protect cells from infection (32, 33). In each of the type I systems, several of the subtype-specific Cas proteins assemble into crRNA-guided surveillance complexes. These complexes find and bind target sequences complementary to the crRNA spacer.

Figure 1

Adaptive immunity in bacteria and archaea is mediated by diverse sets of CRISPR (clustered regularly interspaced short palindromic repeat) loci and their associated genes. (*a*) CRISPR-associated (*cas*) genes (*gray arrows*) encode proteins required for new spacer sequence acquisition (stage 1), CRISPR RNA (crRNA) biogenesis (stage 2), and target interference (stage 3). Each CRISPR locus is flanked by an adenine- and thymine-rich leader sequence (*teal*), followed by a series of direct repeats (*black rectangles*) separated by unique spacer sequences acquired from invading genetic elements (protospacers). Protospacers are flanked by a short motif called the protospacer-adjacent motif (PAM). Long CRISPR transcripts (pre-crRNA) are processed into short crRNAs that guide Cas proteins to invading nucleic acids through complementary base pairing. (*b*) CRISPR loci and their associated genes are remarkably diverse, but phylogenetic analysis performed by Makarova et al. (31) has described three types (I, II, and III) of CRISPR-mediated immune systems, which are further divided into subtypes (e.g., A, B, C). The *cas* genes (*arrows*) are labeled according to the new nomenclature, but some of these genes are still referred to by their previously established names (*labels below the arrows*). Structures for many of the Cas proteins have been determined, and atomic coordinates have been deposited in the Protein Data Bank (PDB). The PDB identification numbers are listed above the arrows (*asterisks* indicate that more than one structure has been determined). In some immune systems, specific Cas proteins (*blue boxes*) assemble into large complexes that include a crRNA. Specific sets of *cas* genes cosegregate with CRISPR loci that have a particular repeat sequence type, which are shown as sequence logos (14). A consensus repeat sequence for CRISPR loci associated with type I-D and type II-B systems has not been reported. The sequence logos for these two subtypes were generated using CRISPR repeats from *Natronomonas pharaonis* DSM 2160, *Haloquadratum walsbyi* C23, *Halorubrum lacusprofundi* ATCC 49239, and *Methanospirillum hungatei* JF-1 for the type I-D system. The type III-B logo was created using CRISPR repeats from *Francisella novicida* U112, *Wolinella succinogenes* DSM 1740, *Francisella cf. novicida* Fx1, *Francisella cf. novicida* 3523, and *Legionella pneumophila* str. Paris. Repeat sequences are diverse even within immune system subtypes, but most have a conserved 3′-terminal GAAA motif (*vertical green box*). Many of the Cas proteins are predicted to be or have been biochemically shown to function as helicases (*black arrows*) or nucleases (*red arrows*), and these are sometimes fused into a single protein (*red and black arrows*). Some of the *cas* genes (*cas1* and *cas2*) are highly conserved (*solid red arrows*), whereas others are specific to certain immune systems (*gray arrows*). In some immune systems, the protospacer sequences selected for integration are flanked by a two- to five-nucleotide PAM (11, 27, 74, 81, 83).


a Type I
Cas1, 2
PAM
Foreign DNA
5′
3′
AAG
TTC
Protospacer
Acquisition
cas3
cas6
cas1
cas2
L
R
S1
S2
S3
cas gene
CRISPR
crRNA-guided surveillance
Spacer
G
C
Cas3
Interference
crRNA biogenesis
crRNA
b Type II
AGAAT
TCTTA
cas9
tracrRNA
Cas9
RNase III
c Type III
PAM
?
Foreign DNA/RNA
cas10
Type III-B

The CRISPR-associated complex for antiviral defense (Cascade) from *E. coli* K12 (type I-E) was the first crRNA-guided surveillance complex described (32, 33). Cascade is a 405-kDa ribonucleoprotein complex composed of 11 subunits of five functionally essential Cas proteins (33). One of the subunits is a CRISPR-specific endoribonuclease (Cas6e; formerly referred to as CasE or Cse3) that cleaves long CRISPR RNA into mature 61-nt (nucleotide) crRNAs (32–35). Cas6e and the crRNA are required for stable assembly of the other Cas proteins (Cse1, Cse2, Cas7, and Cas5). Using cryoelectron microscopy (cryo-EM), researchers recently determined a structure of the Cascade complex (36). This structure provides a detailed description of the subunit organization and explains how the Cas proteins protect the crRNA from degradation while maintaining availability of the crRNA for complementary base pairing with an invading nucleic acid. A similar complex exists in *S. solfataricus* (type I-A), and a crystal structure of Cas7 from this system reveals a crescent-shaped molecule that may cradle the crRNA (37). In the *S. solfataricus* (type I-A) and *E. coli* (type I-E) complexes, Cas7 assembles into a right-handed helix along the ribose-phosphate backbone of crRNA, forming a ribonucleoprotein filament that is morphologically similar to the RecA nucleoprotein filament (36–38). The crRNA guides Cascade to its dsDNA target sequence via an ATP-independent process in which the crRNA base pairs with a complementary DNA strand, displacing the noncomplementary strand to produce an R loop (33). Base pairing in the target-bound complex extends along the length of the crRNA, resulting in a series of short helical segments reminiscent of the base-pairing interaction mediated by the RecA nucleoprotein filament, wherein the DNA is globally underwound and stretched but locally allowed to adopt a B DNA–like conformation (36, 38). This suggests that Cas7 pre-positions crRNA in an underwound and stretched conformation optimal for strand invasion and exchange (transition state stabilization), similar to that described for RecA (37, 38). However, unlike RecA, which subsequently catalyzes DNA repair, target recognition by Cascade may induce a conformational change that recruits Cas3 for destruction of the invading DNA (36).

Large crRNA-guided surveillance complexes have been identified in several type I systems, and low-resolution structures are available for the complexes from *S. solfataricus* (type I-A) (37), *Bacillus halodurans* (type I-C) (39), and *Pseudomonas aeruginosa* (type I-F) (40). Although these structures suggest that type I complexes may share a similar Cas7-based helical platform, each complex is visibly distinct, and differences in their nucleic acid–binding properties have been reported.

Type II CRISPR-Associated Systems

Type II systems have been found only in bacteria (31). These systems consist of only four *cas* genes: *cas9*, *cas1*, *cas2*, and either *csn2* (type II-A) or *cas4* (type II-B) (**Figures 1*b*, 2*b***). The *cas9*

Figure 2

Structural and functional differences among the three CRISPR/Cas types. Protospacers in type I systems (*a*) are flanked by a 5′ PAM, whereas protospacers in type II systems (*b*) are flanked by a 3′ PAM. PAM sequences have not been identified in type III systems (*c*). Protospacers are integrated into the leader (L; *teal arrow*) end of the CRISPR locus, and the repeat sequence is duplicated, maintaining the repeat-spacer-repeat architecture. CRISPR loci are transcribed, and in type I and type III systems, CRISPR-specific endoribonucleases (i.e., Cas6 family proteins) nucleolytically process the long CRISPR RNA. In type II systems, a *trans*-activating crRNA (tracrRNA) hybridizes to each repeat sequence in the CRISPR RNA, and RNase III cleaves these short (~24-nucleotide) duplexes. The type I and type II systems target double-stranded DNA. In type I systems, the crRNA assembles into a multisubunit surveillance complex (e.g., Cascade in type I-E). Target binding induces a conformational change that bends the double-stranded DNA target and promotes R-loop formation (33, 36, 114). Cas3 is a *trans*-acting nuclease that degrades the target (114, 121–124). Target interference in type II systems requires only a single protein (i.e., Cas9) and two RNAs (i.e., crRNA and tracrRNA) (41–43). Type III-A systems are expected to target incoming DNA (44), whereas type III-B systems target single-stranded RNA (23, 45, 46).

gene is a hallmark of this system and encodes a large multifunctional protein that participates in both crRNA biogenesis and in the destruction of invading DNA (41–43). crRNA biogenesis in type II systems is unique in that it requires a *trans*-activating crRNA (tracrRNA). In *Streptococcus pyogenes*, the tracrRNA is encoded upstream and on the opposite strand of the CRISPR/Cas locus (41). Two isoforms of the tracrRNA are expressed (89 and 171 nt), both of which contain a 25-nt stretch that is almost perfectly complementary (one mismatch) to the repeat sequences in the CRISPR. Hybridization between the tracrRNA and the crRNA repeats results in dsRNA that is recognized and cleaved by the cellular (non-Cas) RNase III enzyme. Although a deletion of Cas9 inhibits crRNA biogenesis, its precise role in this process is unclear (41). On the other hand, Jinek et al. (43) recently demonstrated that Cas9-mediated cleavage of target DNA requires both the mature crRNA and the tracrRNA. Cas9 proteins include an HNH nuclease domain, which cleaves the DNA strand complementary to the crRNA guide, and a RuvC-like domain that cleaves the noncomplementary strand (43).

Type III CRISPR-Associated Systems

Two type III systems have been identified (type III-A and type III-B) (31). These systems are most common in archaea, and the type III-B system is found only in conjunction with other CRISPR types. The two type III systems both encode *cas10* and *cas6* genes (**Figure 1*b***). Cas6 is a CRISPR-specific endoribonuclease, and Cas10 may be involved in target interference. Despite these similarities, the two systems appear to target chemically different substrates. The type III-A system of *Staphylococcus epidermidis* targets DNA (44), whereas the type III-B systems in *Pyrococcus furiosus* and *S. solfataricus* cleave target RNA (**Figure 2*b***) (23, 45, 46). This fundamental difference highlights the functional diversity present even within the same CRISPR/Cas type.

IMMUNE SYSTEM ACTIVATION AND REGULATION

Bacteria and archaea perceive and respond to changes in their environment through signaling cascades that often result in transcriptional reprogramming. Genome-wide analysis of the cellular response to phage challenge has been reported for two different model systems (48, 49). Microarray analysis in *Thermus thermophilus* (HB8) demonstrated that some of the *cas* genes are constitutively expressed, and many of these transcripts accumulate during phage infection (48). The cyclic AMP (cAMP) receptor protein (CRP) controls a subset of these *cas* genes (48, 50), and the authors speculate that cAMP may serve as an important signaling molecule to stimulate the immune system. This suggestion is supported by the recent discovery that some of the Cas proteins contain adenylyl cyclase–like domains, which may be involved in cyclic dinucleotide synthesis (51, 52). Cyclic (di-)nucleotides may serve as an alarm signal that binds to and activates transcription factors such as CRP or the Cas proteins Csa3 and Csx1, which contain dinucleotide-like binding domains (53).

In *Streptococcus thermophilus* (DGCC7710), temporal analysis of the immune response to phage challenge was performed using high-throughput protein profiling (49). This analysis revealed dynamic differences in both the host and viral proteome over the course of infection. Similar to *T. thermophilus*, some of the *Streptococcus thermophilus* Cas proteins are constitutively expressed, and many were significantly induced during infection.

Together these studies demonstrate that the CRISPR/Cas systems in *S. thermophilus* and *T. thermophilus* both respond to phage challenge, but the type I-E system appears to be the primary defense system in *T. thermophilus*, whereas the type II system acts as the primary response in *S. thermophilus*. The reason for preferential activation of one immune system type over another is not yet clear, but different immune system types may be differentially

expressed in response to their effectiveness against certain parasites.

In addition to the CRISPR/Cas systems, these two studies indicated that alternative immune systems are also activated in response to phage challenge. In *S. thermophilus*, restriction-modification proteins are upregulated during phage infection (49), and in *T. thermophilus*, the gene encoding an Argonaute protein is upregulated (48). The role of Argonaute proteins in prokaryotes remains unknown, but they may participate in genome defense (30). Collectively, the two studies suggest that phage challenge elicits both innate and adaptive defense systems, though further investigation is required to understand how or if these systems are functionally integrated.

Although pathogen detection and rapid activation of the immune response are critical to surviving an infection, an uncontrolled immune response can be detrimental (i.e., autoimmune reactions). In *E. coli* K-12 and *Salmonella enterica* serovar Typhimurium, the CRISPR loci and *cas* genes are stringently repressed by a histone-like nucleoid protein (H-NS) (25, 54, 55). H-NS is a global transcriptional repressor that preferentially binds to DNA sequences that are AT rich (56, 57). After initially binding to a nucleation site (58, 59), H-NS proteins cooperatively spread along the DNA (60), generating extended nucleoprotein filaments that render larger regions of the affected chromosome inaccessible to RNA polymerase.

The transcriptional activator LeuO mediates derepression of the CRISPR/Cas system. LeuO is a potent antagonist of H-NS that binds upstream of the *casABCDE12* operon and blocks the cooperative spreading of H-NS along the DNA (25, 54, 55, 61). Although we know LeuO (activator) and H-NS (repressor) have opposing roles in immune system regulation, the cellular signals that govern the on/off balance are less well understood. During slow growth conditions, the small molecule guanosine tetraphosphate (ppGpp) induces LeuO (62, 63). ppGpp is an alarmone that functions as a chemical messenger by activating the stringent response (64). However, attempts to induce the CRISPR/Cas systems by amino acid starvation neither increased transcription of the *casABCDE12* operon nor elevated levels of mature crRNA (55, 65).

Regulatory control of the CRISPR/Cas system may not be restricted to LeuO. Numerous studies performed using phylogenetically diverse organisms have shown that the *cas* genes are induced in response to a wide variety of abiotic stimuli (e.g., UV light, ionizing radiation, ionic strength, heat) (66–69), and a recent report by DeLisa and colleagues suggests that misfolded membrane proteins may trigger an envelope stress response that activates CRISPR/Cas expression through the BaeSR two-component signaling system (70). Although the details of immune system activation remain unclear, disturbances at the cell surface may prove to be an important mechanism.

In addition to stress-dependent activation of the *cas* genes, several studies reported basal expression of CRISPR RNA even in nonstress conditions (23, 41, 71–74). This is consistent with a role for CRISPRs as rapid response sentinels that are constitutively ready for defense against previously encountered foreign genetic elements.

THREE STAGES OF CRISPR-MEDIATED IMMUNITY

The mechanism of protection in each of these CRISPR/Cas systems can be divided into three general stages: CRISPR adaptation (i.e., new spacer acquisition), crRNA biogenesis, and crRNA-guided interference (**Figure 1*a***).

CRISPR Adaptation

In 2005, three independent studies reported that spacer sequences within CRISPR loci were often identical to sequences from phages and plasmids (22, 75, 76). Now almost a decade later, the remarkable insight offered by these three computational studies deserves reflection. Pourcel and colleagues (22) reported that the CRISPR loci in *Yersinia pestis* evolve by the polarized addition of new phage-derived spacer

sequences and that new sequence acquisition is accompanied by the duplication of the repeat sequence nearest the leader end of the CRISPR. A complementary report by Bolotin et al. (75) indicated that phage sensitivity in *S. thermophilus* correlated with the number of spacers in the CRISPR locus that were homologous to DNA in the challenging phage. However, they could not detect consensus sequences among the different spacer sequences, but when they aligned 70-bp fragments of the phage DNA—comprising 30 bp of the spacer match and 20 bp of the adjacent phage genome—they discovered a conserved sequence motif in the phage DNA that was located downstream of each spacer match (75). Although the importance of this motif was not realized at the time, these short sequence motifs have become a focal point of current research in CRISPR biology [now called protospacer-adjacent motifs (PAMs); see below]. Finally, Mojica et al. (76) performed a comprehensive analysis of all spacers collected from CRISPR loci in genome sequences available at the time. This analysis indicated that the integration of foreign DNA into CRISPR loci is a widespread phenomenon and that CRISPR transcripts might be central components of a new phage defense mechanism similar to RNA interference in eukaryotes (76).

Phage challenge experiments in *S. thermophilus* have played a pivotal role in our understanding of CRISPR-mediated immunity. *S. thermophilus* is a gram-positive bacteria routinely used in the dairy industry for large-scale production of yogurt and cheese (reviewed in 77). However, bacteriophage infection in industrial fermentors can lead to the lysis of these cultures, resulting in significant economic losses (78). Therefore, the development of phage-resistant strains of bacteria for use in industrial fermentation has been a major focal point of research in the dairy industry (79). In a remarkable collaboration between industry and academia, Barrangou et al. (80) tested the hypothesis that CRISPRs are part of an adaptive immune system by challenging an industrial strain of *S. thermophilus* with two different phages isolated from yogurt samples and then screened these cultures for bacteriophage-resistant mutants. Nine phage-resistant mutants of *S. thermophilus* were isolated, and all of these strains contained between one and four new spacer sequences. Consistent with previous bioinformatic observations, all the new spacer sequences were added in a polarized fashion at the leader end of the CRISPR locus, and the addition of a repeat sequence accompanied each new spacer. Naturally acquired spacer sequences were derived from either strand of the invading DNA (sense and antisense), and the number of new phage-derived spacers correlated with the degree of phage resistance. Importantly, single-nucleotide polymorphisms between the spacer and the phage sequence did not provide resistance, suggesting that the sequence of the spacer was critical for protection. In line with this observation, Barrangou and colleagues (80) demonstrated that phage resistance could be augmented or erased through insertion or deletion (respectively) of phage-targeting spacer sequences in the CRISPR locus.

The rapid acquisition of new spacer sequences in *S. thermophilus*, along with well-maintained collections of phages, has made this a powerful model system for studying CRISPR adaptation. By performing successive phage challenges, Deveau et al. (81) demonstrated that the iterative addition of spacers could expand the repertoire of phage protection. However, these loci do not grow ad infinitum, and the occasional loss of repeat-spacer units has been observed. These deletions usually occur toward the trailer end (opposite the leader) of the CRISPR locus, possibly supporting preferential elimination of outdated spacers that target ancient phages or plasmids while maintaining the more contemporary arsenal of spacers at the leader end.

The CRISPR/Cas machinery appears to target specific sequences for integration into the CRISPR locus (**Figure 3**). Sequences in foreign DNA selected for integration are called protospacers, and these sequences are often flanked by a short sequence motif, commonly referred to as the PAM (**Figures 1–3**) (75, 81,

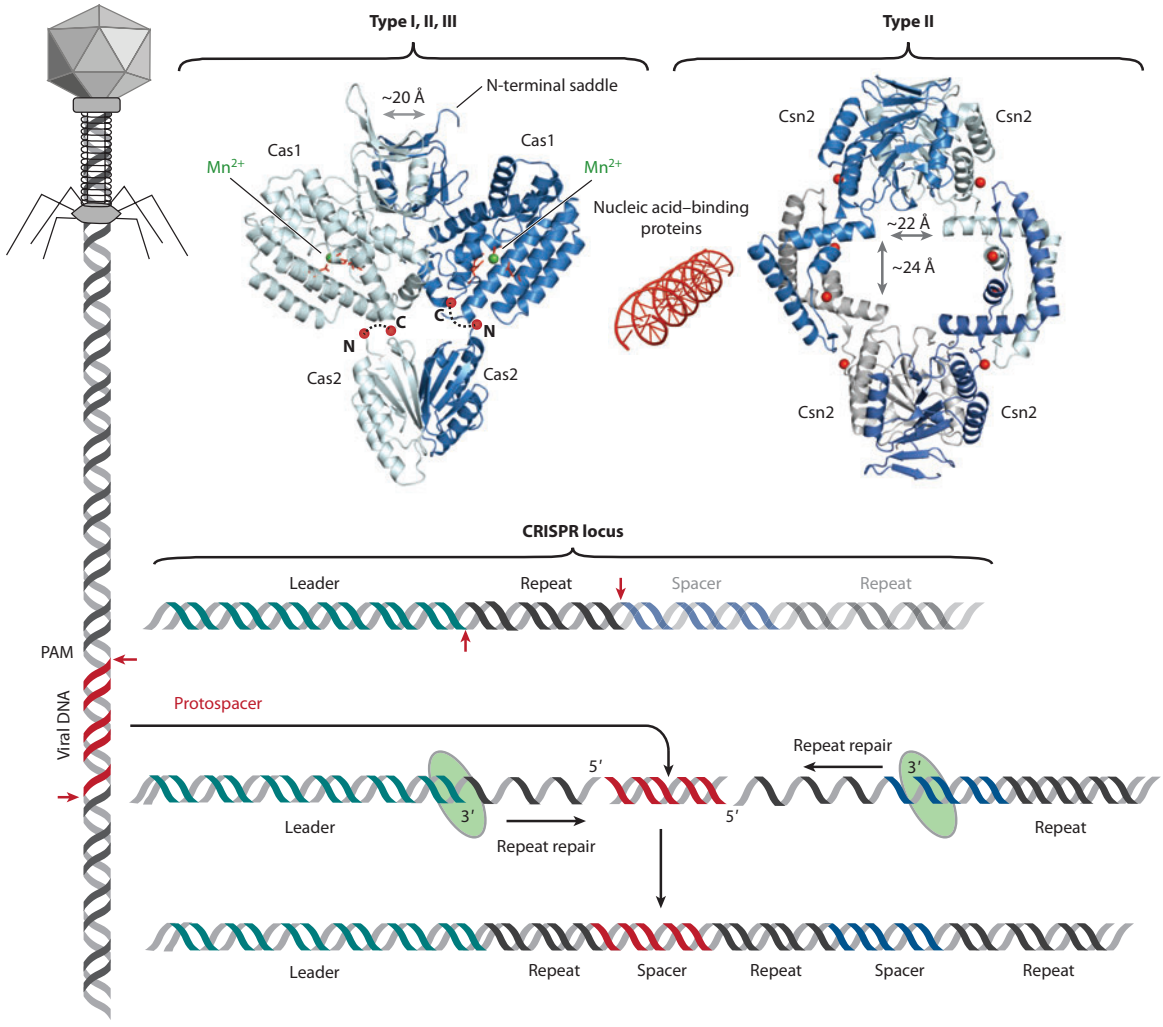


Figure 3

Protein and nucleic acid requirements for new sequence acquisition. (*Top*) Cas1 and Cas2 proteins are required for new sequence acquisition in all CRISPR/Cas systems (27). In some immune systems, the C-terminal domain of Cas1 is fused to the N-terminal domain of Cas2 (*red dots* and *dashed lines*) (68). Cas1 is a metal-dependent DNase that forms a stable homodimer in which the two molecules (*cyan* and *blue*) are related by a pseudo-twofold axis of symmetry (PBD ID: 3GOD) (93, 95). This organization creates a saddle-like structure in the N-terminal domain of Cas1 that can be modeled onto double-stranded DNA without steric clashing (13). The metal ion (Mn^{2+}, *green sphere*) in the C-terminal domain of each Cas1 subunit is surrounded by a cluster of basic residues that have also been implicated in non-sequence-specific DNA binding (13, 93, 95). Cas2 proteins have a ferredoxin-like fold (PDB ID: 3OQ2), and two protomers (*cyan* and *blue*) assemble into a stable homodimer (94, 158), reminiscent of the duplicated ferredoxin-like fold found in some CRISPR-specific endoribonucleases. Cas2 proteins have been implicated as metal-ion-dependent nucleases in some systems (94). Genetic experiments in *Streptococcus thermophilus* indicate that Csn2 is required for new sequence integration in type II-A systems (80). Csn2 is a Ca^{2+}- (*red spheres*) dependent DNA-binding protein that assembles into a homotetrameric ring with a positively charged inner pore large enough to accommodate double-stranded DNA (PDB ID: 3S5U) (88–90). (*Bottom*) Foreign DNA selected for integration into the CRISPR locus is referred to as a protospacer (*red*) (81). Protospacers are not selected for integration at random (83). In type I and type II systems, protospacers are flanked by a short motif called the PAM. New sequence acquisition requires Cas1 and Cas2 as well as a leader sequence and an adjacent repeat sequence (27). The precise mechanism of integration remains undetermined, but the coordinated cleavage of the foreign DNA (*red arrows*) and integration of the protospacer into the leader end of the CRISPR occur via a mechanism that duplicates the leader-proximal repeat sequence (27) and may require cellular DNA repair proteins (*green ovals*) (13, 29, 93, 95).

82). Interestingly, the sequence and location of the PAM vary according to the CRISPR/Cas type (**Figures 1*b*** and **2*a*,*b***) (83). The variability of this motif was initially observed by comparing PAM sequences among different CRISPR loci in *S. thermophilus* (81, 82). Similarly, phage challenge experiments in *Streptococcus mutans* revealed that the PAM for one CRISPR locus was 3′ of the protospacer, whereas another CRISPR locus had a different motif located 5′ of the protospacer (86). PAM sequences also varied among different CRISPR loci in members of the Sulfolobales (24). In 2009 Mojica et al. (83) performed a comprehensive analysis of PAM sequences that revealed distinct PAMs that correspond to specific CRISPR/Cas subtypes. This suggests that different CRISPR loci evolve using different selection criteria that may be specified by the Cas proteins.

Although the mechanism of spacer integration is still unknown, genetic studies in *S. thermophilus* and *E. coli* have indicated that several Cas proteins are involved in the process (**Figure 3**). Mutational analysis of the *cas* genes in the type II-A system of *S. thermophilus* has demonstrated that *csn2* is required for new spacer sequence acquisition (42, 80, 87). The precise role of this protein remains unclear, but biochemical and structural studies have shown Csn2 is a calcium-dependent dsDNA-binding protein that assembles into a tetrameric ring with a positively charged inner pore (~26 Å wide) large enough to accommodate dsDNA, but alternative binding modes are also being considered (88–90). In type II-B CRISPR systems, the *cas4* gene replaces *csn2* (28, 29, 31). Cas4 contains a RecB-like nuclease domain that may be involved in CRISPR adaptation (21, 28, 29, 31, 47), and the conserved synteny between *csn2* and *cas4* suggests that they may have orthologous functions. Neither *csn2* nor *cas4* is conserved in all CRISPR/Cas systems, suggesting that either the mechanism of adaptation in type II systems is distinct from that in the other types, or that functional orthologs of these proteins exist in the other systems.

Until recently, the type II CRISPR/Cas systems in *Streptococcus* spp. have been the primary model for studying new spacer acquisition. Phage challenge experiments performed in other pure culture systems have failed to provoke CRISPR adaptation. However, a series of recent reports demonstrated that CRISPR loci associated with the type I-E and type I-F systems can be activated (27, 84, 85, 91, 92). Cas1 and Cas2 are conserved nucleases involved in integration (27, 29, 93–95). Astonishingly, overexpression of *cas1* and *cas2* is sufficient to result in the addition of new spacer-repeat units at the leader end of an endogenous CRISPR locus in the *E. coli* BL21 (DE3) chromosome (27). The precise mechanism for leader-end recognition remains unknown, but mutations in the leader sequence that block transcription do not interfere with integration (i.e., transcription of the CRISPR locus is not required for integration). In contrast, mutations in the first 60 nt of the leader abolish integration. The first 60 nt of the leader sequence are essential, but integration does not occur without at least one repeat sequence. Although the origin of new spacer sequences is well established, the mechanism for generating new repeats has not yet been described. By introducing single-nucleotide mutations into repeat sequences, Yosef and colleagues (27) demonstrated that only mutations in the first repeat are propagated during the integration process, proving that the leader-proximal repeat is the template for subsequent generations of repeat sequence.

In the plasmid transformation experiments performed by Yosef and colleagues (27), most of the spacers integrated into the CRISPR derived from plasmid DNA, and all of the new inserts derived from regions of DNA with an adjacent PAM sequence. This suggests that Cas1 and/or Cas2 are capable of recognizing the 3-nt PAM sequence. Although the observed bias for acquiring plasmid-derived spacers may reflect a mechanism for selecting nonself (plasmid) DNA, acquisition of spacers from the *E. coli* chromosome may also kill the cell and thus reduce the apparent frequency of self-acquisition events, a process denoted as CRISPR-mediated autoimmunity (96). The potential toxicity of acquiring spacers from

cellular DNA may explain why H-NS carefully represses the transcription of the *cas* genes in this system (25, 26, 55). Systems with additional cofactors that guide Cas1 and Cas2 to foreign targets and prevent self-targeting may not require such stringent regulation. Although these cofactors have not been identified, it is tempting to speculate that Csn2 or Cas4 may offer an additional layer of self-/nonself-discrimination that permits the constitutive expression of the integration machinery in these systems.

In addition to Cas1 and Cas2, the type I-E CRISPR system in *E. coli* includes six other Cas proteins. Although these proteins are not required for new sequence acquisition (27, 32), the pattern of new sequence acquisition changes when they are included (91, 92). Most notably, in an *E. coli* system that includes a full complement of all eight (type I-E family) *cas* genes, the CRISPR locus often expands by integrating multiple new spacer sequences (91, 92). The addition of the first spacer accelerates acquisition of subsequent spacers, and all of the spacers derive from the same strand of DNA (91, 92). This strand bias is established by the first new sequences added to the CRISPR and is not observed in the minimal system, which includes only Cas1 and Cas2 (91, 92). This suggests that the Cas proteins previously implicated in target interference (i.e., Cascade and Cas3) may play supporting roles in the process of new sequence acquisition by localizing the CRISPR/Cas machinery to invading DNA (91, 92). The level of bacteriophage resistance increases with the number of target-specific spacers; thus, a mechanism for rapid expansion of the CRISPR locus in response to a specific signal may limit opportunities for phage-escape mutants to evolve.

The recently established integration systems in *E. coli* permit reevaluation of the PAM sequence. The consensus sequence for type I-E was originally predicted to be 5′-AWG (83), which is consistent with the 5′-AAG PAM found in association with the majority (~80%) of all newly acquired spacers (91, 92). However, on occasion, protospacers with noncanonical PAMs were acquired (i.e., AAA and AAT). The PAM has been defined as the motif located adjacent to the protospacer, but the acquisition of protospacers with an "A" (5′-AAA) or "T" (5′-AAT) in the last position resulted in a concomitant change in the last nucleotide of the repeat sequence. This indicates that the last nucleotide of the PAM is actually part of the protospacer. The sequence and the location of the PAM vary among the different systems, but this observation indicates that the PAM is not restricted to the region adjacent to the protospacer and that the PAM can also be part of the protospacer. Thus the PAM is not always a protospacer-adjacent motif but rather a protospacer-associated motif.

The importance of the PAM goes beyond protospacer selection. Plasmids or phages that have a single mutation in the PAM are no longer sensitive to CRISPR-mediated interference, even when the spacer and protospacer sequences are 100% complementary (42, 81, 97, 98). Although the PAM sequences play a critical role in new sequence acquisition and target interference in type I and type II systems, these sequences appear to be absent in type III systems (31, 83). This implies that type III systems may rely on a distinct mechanism for new sequence selection, but active integration systems have not been reported for type III systems, and the mechanism of protospacer selection remains untested.

CRISPR RNA Biogenesis

The maturation of crRNAs is critical to the activation of all CRISPR/Cas immune systems and involves at least two distinct steps. CRISPR loci are initially transcribed as long precursor crRNAs (pre-crRNAs) from a promoter sequence in the leader (**Figure 2**). Subtype-specific enzymes then process these pre-crRNAs into mature crRNA species (**Figure 4**). Pre-crRNA processing in type I and type III systems involves a diverse family of CRISPR-specific endoribonucleases that exclusively interact with the repeat sequences from their associated CRISPR loci (34, 35, 99–104). In contrast, type II systems rely on a

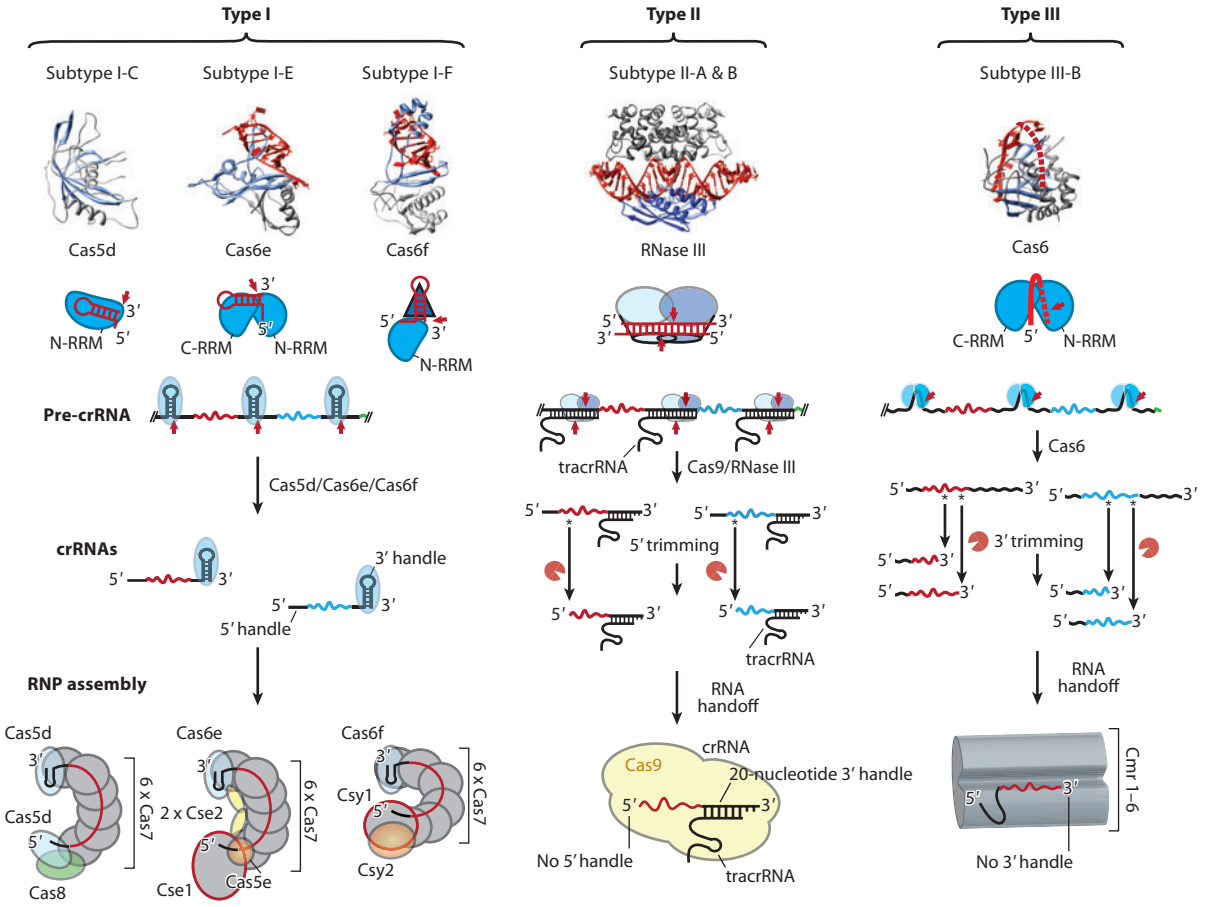


Figure 4

Pre-crRNA processing and assembly of crRNA-guided surveillance complexes. Pre-crRNA processing is essential for activating crRNA-guided interference in all CRISPR/Cas systems, but the mechanisms of RNA recognition and cleavage are diverse. In type I and type III systems, pre-crRNA processing relies on a diverse family of CRISPR-specific endoribonucleases that have an N-terminal RNA recognition motif (N-RRM), a C-terminal RNA recognition motif (C-RRM), or both (*blue half circles*). Crystal structures of these proteins alone (Cas5d, PDB ID: 4F3M) or in complex with their cognate crRNAs (Cas6e, PDB ID: 2Y8W; Cas6f, PDB ID: 2XLK) reveal unique tertiary folds for each of these proteins and distinct mechanisms for RNA recognition (34, 35, 39, 53, 102–104). Cas6e and Cas6f interact exclusively with their respective crRNA substrates by making sequence- and structure-specific interactions in the major groove of each stem-loop structure. In contrast, the repeat sequences in type III-B systems are predicted to be unstructured, and the 5′ end of the single-stranded RNA repeat sequence is wedged in a positively charged cleft of Cas6 (PDB ID: 3PKM) created by opposing β-sheets on the N-RRM and C-RRM (108). All CRISPR-specific endoribonucleases (type I, *blue ovals*; type III, *blue half circles*) cleave within each repeat sequence, generating mature crRNAs consisting of a spacer sequence flanked by 8 to 10 nucleotides of the repeat sequence on the 5′ end (known as the 5′ handle) and ~20 nucleotides of the remaining repeat sequence on the 3′ end (referred to as the 3′ handle) (34, 35, 39, 53, 102–104). In type I systems, the CRISPR-specific endoribonuclease and the mature crRNA are assembled into large ribonucleoprotein complexes that serve as crRNA-guided surveillance complexes. Pre-crRNA processing in type II systems relies on a host-encoded RNase III enzyme (PDB ID: 2EZ6) and a tracrRNA (41). The 5′ end of the crRNA is trimmed (*black asterisk* and *arrow*) (41), and both RNAs are required for targeting by Cas9 (43). Trimming in type III systems occurs at the 3′ end, and Cas6 does not retain the mature crRNA in type III systems. In type III systems, the mature crRNA is handed off to a Cas protein complex [e.g., type III-B is called the Cas repeat-associated mysterious proteins module (Cmr) complex].

completely different mechanism that involves Cas9 recognition and cleavage of dsRNA repeats by a host-encoded RNase III (41).

In 2008, Brouns et al. (32) identified a CRISPR-specific endoribonuclease called Cas6e (formerly CasE or Cse3) responsible for pre-crRNA processing in *E. coli* (type I-E). Cas6e is a member of a large family of extremely diverse proteins referred to as RAMPs (repeat-associated mysterious proteins) (28, 29, 105). All RAMP proteins contain at least one RNA recognition motif (RRM) (also referred to as a ferredoxin-like fold) and a conserved glycine-rich loop (G loop). RRMs consist of a conserved $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ arrangement in which the β-strands are arranged in a four-stranded antiparallel β-sheet and the two helices are packed together on one face of the sheet (106). This fold is found in a wide variety of RNA-binding proteins that usually bind RNA using conserved residues positioned along the open face of the β-sheet. Crystal structures of the Cas6e protein reveal a two-domain architecture consisting of one N-terminal and one C-terminal RRM (**Figure 4**) (34, 35, 107). A short proline-rich linker connects the two domains, and the β-sheets in each RRM face one another, creating a V-shaped groove that runs along one surface of the protein. This cleft was initially predicted to be the RNA-binding surface (107), but cocrystal structures of Cas6e bound to its crRNA substrate reveal a noncanonical binding mechanism that involves a unique combination of sequence- and structure-specific interactions primarily located on the opposite face of the protein (34, 35). The repeat sequence of the *E. coli* CRISPR is partially palindromic, producing an RNA with a stable 7-nt stem capped by a GCGU tetraloop. A positively charged β-hairpin in the C-terminal domain of Cas6e interacts with the major groove of the RNA duplex and positions the phosphate backbone of the 3′ strand of the crRNA stem along a positively charged cleft that runs the length of the protein. RNA binding induces a conformational change that disrupts the bottom base pair of the stem and positions the scissile phosphate in an extended conformation in the enzyme active site (35). The cleavage mechanism is independent of metal ions and occurs at the base of the stem, generating a mature crRNA with a 5′ hydroxyl and a 2′,3′-cyclic phosphate (33). The mature crRNAs (~61 nt) consist of a 32-nt spacer flanked by 8 nt of the repeat sequence on the 5′ end (known as the 5′ handle) and 21 nt of the remaining repeat sequence on the 3′ end. Cas6e remains bound to the 3′ stem-loop and may serve as a nucleation point for assembly of a large surveillance complex called Cascade (**Figure 4**), which is required for target silencing in the next stage of the immune system (32, 33, 36).

Crystal structures of the Cas6 protein from *P. furiosus* (types III-B, I-A) reveal the same duplicated ferredoxin-like fold observed in the Cas6e protein (100, 108, 109). However, biochemical and structural studies reveal a distinct mechanism for RNA recognition that involves the more canonical (β-sheet) face of the protein. The repeat sequences in CRISPR loci associated with this system are predicted to be unstructured, and the 5′ end of the single-stranded RNA repeat sequence is wedged in a positively charged cleft created by opposing β-sheets in each RRM (**Figure 4**) (108). Although the crRNA is disordered in the enzyme active site, biochemical studies have shown that cleavage occurs specifically between an AA dinucleotide located 8 nt upstream of the spacer sequence, generating a 5′ handle similar in length but different in sequence compared with the crRNA generated in the *E. coli* system (99). Unlike pre-crRNA processing in the type I-E system, Cas6-mediated cleavage in this system results in a 69-nt crRNA intermediate (sometimes referred to as the 1x intermediate) that is further processed by nucleolytic 3′-end trimming (**Figure 4**). Three-prime-end trimming in *P. furiosus* results in two distinct populations of mature crRNAs (39 and 45 nt long) that lack the 3′-repeat sequence (23, 45, 99, 100, 110). In *E. coli*, Cas6e remains associated with the 3′-repeat sequence, but in *P. furiosus*, the 3′ repeat is removed, and the crRNA is loaded into a large protein complex composed of subtype-specific Cas proteins

called Cmr (Cas module RAMP) proteins (23, 45). Importantly, loading the crRNA into the Cmr complex requires the 5′ handle (23).

Pre-crRNA processing in *S. epidermidis* (type III-A) occurs via a similar mechanism that initially involves Cas6-mediated cleavage of the pre-crRNA followed by 3′-end trimming (**Figure 4**) (101). 3′-end trimming results in two crRNA species that are 37 and 43 nt long, but the mature species contain a 3′-hydroxyl group rather than a 2′–3′-cyclic phosphate characteristic of the 1x intermediate (101). It is not clear why 3′-end trimming in both type III systems produces mature crRNAs with two different lengths, though it is interesting that the two crRNA populations within each system differ in length by 6 nt. This may indicate a common structural explanation for the ruler mechanisms that define these two crRNA species.

Crystal structures for CRISPR-specific endoribonucleases from two other immune systems have been determined. CRISPR loci associated with the type I-C and type I-F systems contain repeat sequences that are partially palindromic, but the proteins that bind these repeats are structurally distinct. The Cas6f protein (also known as Csy4) from *P. aeruginosa* contains an N-terminal RRM that is structurally similar to Cas6e and Cas6, but this domain does not interact with the crRNA. Instead, an arginine-rich α-helix in the structurally unique C-terminal domain of Cas6f inserts into the major groove of the crRNA duplex, forming a complex network of hydrogen-bonding interactions that are highly sensitive to the helical geometry of the crRNA substrate (102–104). This shape-based recognition mechanism is reminiscent of the so-called arginine fork, which investigators first used to describe the interaction between the HIV-1 Tat protein and the stem of the transactivation response RNA (111). Binding by the arginine-rich helix positions the base of the crRNA stem for sequence-specific hydrogen-bonding contacts at the base of the RNA major groove. These contacts position the scissile phosphate of the crRNA in an unusual enzyme active site, wherein Ser148 and Tyr176 interact with the 2′ hydroxyl via hydrogen bonding and restrain the ribose ring of the terminal nucleotide in a C2′-endo conformation, resulting in a locally extended RNA backbone required for in-line attack (112). In this proposed mechanism, a conserved histidine abstracts a proton from the pinned 2′ hydroxyl for nucleophilic attack of the adjacent scissile phosphate, resulting in a 2′,3′-cyclic phosphate that hydrolysis by a water molecule may resolve to a 3′ phosphate (103).

The only type I system that does not contain a Cas6-like protein is type I-C (**Figure 1*b***) (31). To identify the pre-crRNA processing enzyme in this system, Nam et al. (39) overexpressed the Cas proteins from *B. halodurans* (strain C-2) and incubated the purified proteins with a pre-crRNA repeat sequence from this system. The Cas5d protein specifically cleaves the crRNA repeat at the 3′ base of the stem-loop, generating a crRNA species with an 11-nt 5′ handle and 21-nt stem-loop at the 3′ end. A crystal structure of the Cas5d proteins reveals an N-terminal ferredoxin-like fold extended by the addition of two β-strands, creating a $\beta_1\alpha_1\beta_2\beta_3\beta_4\beta_5\alpha_2\beta_6$ architecture that differs from the canonical $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_6$ architecture of this fold (39). Extensive mutational analysis of the crRNA reveals that the protein primarily interacts with nucleotides near the base of the crRNA stem, but these interactions are weak, and Cas5d does not stably associate with its crRNA substrate in vitro. Although this weak association has frustrated cocrystallization efforts, the protein and crRNA do assemble with other subtype-specific Cas proteins into a complex that resembles the surveillance machines in other type I systems (**Figure 4**) (39). Interestingly, Cas5-like proteins are found in other type I systems (**Figure 1*b***), but sequence similarities within this family are extremely low, and the other Cas5-like proteins may not possess a similar enzymatic activity.

In contrast to crRNA processing in type I and type III systems, pre-crRNA processing in the type II CRISPR/Cas systems relies on a completely different mechanism (41). Differential deep sequencing of RNAs isolated

from *Streptococcus pyogenes* has uncovered two isoforms of a tracrRNA that contain a 24-nt sequence complementary to the repeat regions of the pre-crRNA. The cellular RNase III enzyme processes short duplexes created by the tracrRNAs and the repeat sequences, creating a unique intermediate species consisting of the crRNA and the 3′ portion of the tracrRNA (41, 43). The crRNA in this hybrid represents a 1x intermediate consisting of a spacer sequence flanked by 13 nt of the repeat on the 5′ end and 21 nt on the 3′ end. Like the 1x intermediate for type III systems, this crRNA undergoes further processing. However, nucleolytic trimming in type II systems occurs at the 5′ end, removing the entire 5′ handle and 6 nt of the 5′ spacer sequence. In vivo processing of crRNAs requires Cas9 (formerly Csn1), which may be involved in recruiting RNase III to the crRNA duplexes or in 5′-end processing. Although the role of Cas9 in crRNA biogenesis remains uncertain, studies recently demonstrated its role in target interference (42, 43).

Finding Your Foe: Target Surveillance and Destruction

Nucleobases provide the necessary molecular structure for hydrogen bonding between complementary strands of DNA or RNA, and this specificity provides a powerful mechanism for molecular recognition. All crRNAs associate with Cas proteins to form large CRISPR-associated ribonucleoprotein complexes, but how do these surveillance complexes find short target sequences complementary to the crRNA in a crowded intracellular environment packed with gigabases of distracting (nontarget RNA and DNA) nucleic acid? Nearly every nucleic acid–binding protein that has a specific recognition sequence faces a similar search problem, and the mechanisms of target finding have been the focus of intense investigation for the past several decades (reviewed in 113). These studies indicate that many DNA-binding proteins are capable of locating their cognate binding sites with much higher levels of efficiency than would be predicted by a process based on purely stochastic collisions. Some proteins accelerate target finding by first localizing on DNA through nonsequence-specific binding events mediated by electrostatic attraction to the negatively charged sugar-phosphate backbone. This initial association is often followed by intramolecular translocations, otherwise known as one-dimensional (1D) sliding. In contrast to the directional motion of energy-consuming motor proteins (i.e., polymerases, helicases, mismatch repair enzymes, and type I restriction enzymes), DNA sliding is an energy-independent process driven by thermal diffusion. Thus, each translocation event has an equal probability of moving along the DNA contour in either direction. However, the initial DNA-binding event is an unguided process, and a search restricted to 1D diffusion may oversample regions of the DNA that do not include a target sequence. Thus, a sequence surveillance process that includes a 3D component may accelerate target finding.

A considerable foundation for understanding the mechanisms of target site location has been established, but there are aspects of the search that may be unique to CRISPR-associated surveillance systems. Target recognition by all CRISPR systems involves hybridization of the crRNA-spacer sequence with a complementary nucleic acid target. This presents a potential problem for CRISPR systems that target dsDNA (type I and II) or for RNA-targeting systems (type III-B) in which secondary structures of the RNA may occlude the target binding sites. The crRNA-guided surveillance complex from *E. coli*, called Cascade, preferentially binds to long dsDNA (plasmid or phage DNA) that is negatively supercoiled (114). Negative supercoiling compacts dsDNA, thus sequences separated by long distances along the contour of the DNA can be positioned in close proximity in 3D space, dramatically accelerating the search process for some dsDNA-binding proteins (115). Furthermore, negative supercoiling introduces torsional tension that facilitates strand separation. Westra et al. (114) recently showed that negative supercoiling provides approximately

half of the energy ($\Delta G_{sc} \approx 90$ kJ/mol) required for separating a 32-nt stretch of dsDNA. However, genetic and biochemical studies suggest that unwinding of the entire 32-nt region may not be necessary for target binding (98). The first 8 nt of the crRNA spacer are most important for target binding (36, 40, 98). This region of the crRNA spacer is referred to as the seed sequence, and single-nucleotide mutations in target sequences complementary to the seed result in major binding defects (98). In contrast, multiple mismatches in the target at nonseed locations are still bound with high affinity and maintain their effectiveness during phage challenge (98). This suggests that initial target binding may require unwinding short regions of the dsDNA target (<8 bp), reducing the cost of unwinding to a fraction of that estimated for a 32-nt stretch. However, a scanning mechanism that relies solely on local unwinding (even short regions) for target recognition may be energetically prohibitive.

A short sequence motif called the PAM is required for efficient target binding (98). The integration machinery initially recognizes the PAM during new sequence acquisition (27), and this same motif is also required for target interference (42, 81, 98). Although the proteins required for foreign DNA acquisition (i.e., Cas1 and Cas2) appear to be distinct from those involved in target interference, the components associated with these two stages of immunity have coevolved to efficiently recognize targets that have a PAM. Cascade binds dsDNA nonspecifically with low affinity (33, 114, 116), consistent with a target-finding mechanism that involves DNA sliding. However, unlike many DNA-binding proteins, which can locate specific sequences using a direct readout (i.e., hydrogen bonding between specific amino acid side chains with the exposed edge of nucleobases in the major groove), crRNA-guided detection requires strand separation for base pairing. The PAM sequence is not complementary to the crRNA, and thus protein interactions likely mediate recognition of this sequence. The Cse1 subunit (also known as CasA) of Cascade is required for binding nonspecific DNA and for sequence-specific interactions, suggesting that this subunit may be involved in engaging DNA for PAM surveillance (116). Although high-resolution crystal structures of the Cse1 protein do not reveal an obvious DNA binding site (116, 117), when docked into the cryo-EM reconstruction of the Cascade complex, they reveal a short disordered loop in Cse1 that appears to be within reach of the target PAM (116, 117), and chemical probing confirms the interaction with the PAM (116). Residues within this loop are important for nonspecific DNA binding by Cascade, suggesting that the loop is important for PAM scanning via a mechanism that may not require strand separation. Detection of the PAM requires a phenylalanine (Phe129), reminiscent of mechanisms observed in the type II restriction endonuclease HinP1I and the DNA repair enzyme MutM (116, 118, 119). In these systems, phenylalanine intercalates into the major groove, causing a local distortion of the B-form helix and a separation of the two strands (118, 119). A similar distortion of the DNA by Cse1 may provide a possible mechanism for initiating crRNA-guided strand invasion.

Hybridization of the crRNA with the target strand generates an R loop and triggers a conformational change in the Cascade complex that coincides with bending of the target DNA (33, 40, 114, 120). However, Cascade binding to the target is not sufficient for target destruction (32, 114). Cas3 is recruited to the target-bound Cascade complex, and the displaced R loop is degraded by a mechanism that relies on conserved residues in the HD nuclease domain (121–124). In many of the type I immune systems (subtypes C, E, and F), the HD domain is fused to a superfamily II helicase domain (29, 31). Nicking of the nontarget strand requires only the HD domain, but processive degradation of the DNA target relies on both the metal-dependent HD domain and the ATP-dependent helicase domain (120, 124).

Unlike type I systems, target interference in type II systems requires only a single protein called Cas9 (formerly Csn1 or Cas7). Barrangou et al. (80) originally demonstrated the importance of Cas9 in target interference in 2007,

but the mechanism of target silencing remained unclear until 2010, when Garneau et al. (42) showed that both strands of the target DNA were cleaved at a specific site within the protospacer, producing a blunt-end cleavage product. Cas9 is a large, multidomain protein containing two predicted nuclease domains (29, 31). Biochemical studies have demonstrated that the Cas9 HNH nuclease domain cleaves the target strand, whereas a Cas9 RuvC-like domain is responsible for cleavage of the noncomplementary strand (43). However, the crRNA is not sufficient for guiding Cas9 to the appropriate target sequence. Pre-crRNA processing in type II systems produces a heterodimeric RNA consisting of a crRNA and a portion of the tracrRNA. Cas9 targeting requires both RNAs for target destruction (43).

The PAM plays an essential role in new sequence integration and in target interference in both type I and type II immune systems, but PAM sequences have not been detected in type III systems. This suggests that the mechanisms for new sequence integration and target interference may be fundamentally different in these systems. The type III immune systems are divided into two different subtypes (III-A and III-B). The type III-A system consists of five subtype-specific *cas* genes, called *csm* genes, which have been implicated in crRNA-guided destruction of foreign DNA (44), whereas type III-B systems consist of subtype-specific *cas* genes called *cmr* genes that are involved in the destruction of target RNA (45, 46, 51). DNA targeting by the type III-A system in *S. epidermidis* does not require a specific PAM sequence, but sequences complementary to both the crRNA spacer sequence and the 5′ handle of the crRNA are not targeted by this system (125). This exclusion mechanism prevents the immune system from targeting spacer sequences in the host CRISPR locus, which are complementary to the crRNA.

In contrast to DNA targeting systems, which must have mechanisms to prevent crRNA-guided self-targeting of the CRISPR loci, RNA targeting systems may not need to make this distinction. Most CRISPR loci are transcribed in one direction and thus do not generate complementary RNA targets. The type III-B systems in both *P. furiosus* and *S. solfataricus* target RNA (23, 45, 46). RNA targeting relies on a large ribonucleoprotein assembly called the Cmr complex. The three-dimensional shape of this complex was recently determined by electron microscopy (46), and high-resolution structures are available for Cmr2 (also known as Cas10) (51, 52), Cmr5 (126), and Cmr7 (46). The Cmr2 protein contains an HD domain that was predicted to be responsible for target cleavage. However, recent biochemical studies have shown that this domain is dispensable for target cleavage in *P. furiosus* (51), and crystal structures of the Cmr2 protein reveal a duplicated adenylyl cyclase domain (51, 52), which may generate modified (di)nucleotides as a signaling molecule, though it is unlikely that Cmr2 serves as the target slicer in this system.

RNA targeting systems may be uniquely capable of providing protection from phages with RNA genomes. An RNA virus that infects hyperthermophilic archaea has been reported and several spacer sequences that are identical to the viral genome have been identified in CRISPR loci from *Sulfolobus* (127). This suggests that type III-B systems may be capable of acquiring resistance to RNA-based viruses, but a mechanism for integrating a DNA copy of the RNA genome into the CRISPR locus has not been demonstrated.

ECOLOGICAL IMPLICATIONS OF ADAPTIVE IMMUNITY

The ecological dynamics of phage and bacteria populations have been studied for decades using a combination of experimental methods that integrate observations from environmental sampling and direct experimentation into theoretical models (128–130). The ecological models take into account numerous parameters, including mutation rates and horizontal gene transfer (HGT), in both the prokaryotic host and phage, aiming to estimate cycles of coevolution and ecological trends. However, the discovery of CRISPR-mediated adaptive

immune systems adds another dimension to these already complicated models. Studies sampling real microbial populations and their CRISPRs over defined timelines have begun to shed light on the ecological implications of adaptive immunity in various ecosystems, including acid mines (131), hot springs (132, 133), the human body (134–136), and the ocean (137). As a result, new models are being developed to incorporate the CRISPR paradigm into ecological and evolutionary interactions (138–142).

Sampling of CRISPRs in Microbial Communities

A pioneering analysis of host-CRISPR-phage interactions at the ecosystem level was performed by the Banfield group (131, 143), who studied microbial communities inhabiting an acidophilic biofilm growing in an acid mine drainage. This microbiome, composed of only a handful of dominant bacterial and archaeal species, was studied by community DNA sequencing (metagenomics), a technique that pools and sequences the DNA of the entire community. Hundreds of metagenomic sequence reads derive from a single CRISPR locus in *Leptospirillum group II*, one of the dominant species in the studied biofilm. Interestingly, the leader end of this CRISPR locus showed extreme heterogeneity, suggesting that no two individual bacteria sampled in that study share the exact same spacer content (143). Moreover, the spacer content of the community significantly changed between two samples taken 5 months apart (131), suggesting rapid evolution of CRISPR loci on a timescale of months, leading to high rates of immune heterogeneity in the population.

Metagenomics has become a popular technique for studying microbial ecosystems that cannot be easily cultivated in laboratory conditions (144). Following shotgun sequencing of the metagenome, overlapping sequence reads are assembled into larger sequences (contigs), which represent DNA sequences from organisms in the community. One of the challenges in these studies is to differentiate between bacterial DNA and the phage DNA associated with a specific bacterial community. In this respect, CRISPR spacers can be used to clarify this ambiguity. DNA contigs that show high similarity to a CRISPR spacer (but not to the flanking repeats) represent phage or plasmid DNA targeted by the CRISPR. Stern et al. (135) recently employed this concept to study phages associated with the human gut microbiome. These researchers reconstructed the CRISPR content of metagenomic gut samples from 124 European individuals (145) and used more than 50,000 retrieved CRISPR spacers to identify almost 1,000 phages associated with human gut bacteria (135). The study revealed a surprisingly high degree of phage sharing among different human individuals, an unexpected finding in light of the extreme phage diversity observed in other ecosystems.

CRISPR spacers not only identify phage DNA but also provide a genetic link between specific microbes and the phages they have encountered. Indeed, analyses of CRISPR genotypes from acid mine drainages and human gut microbiomes have identified the bacterial hosts for newly identified phages, allowing analyses of phage-host distributions across multiple time points and samples (131, 135). Moreover, CRISPR loci also provide a chronological record of infections, with the most recent infection represented by the most leader-proximal spacer. This historical record of cellular infection opens a window to past phage-bacteria interactions in natural ecosystems (141).

Mathematical Modeling

Several mathematical models have been developed in an attempt to assess the implications of CRISPR immunity on phage and bacteria population dynamics (138–142). These models incorporate experimental parameters derived from metagenomic samplings and studies of single species into a simulated ecosystem in which microbes and phages compete. Although most models cannot simulate the entire complexity of the interactions between phages and

bacteria, they do provide insight and testable hypotheses on the ecology of phage and bacterial communities. For example, several models predict that long-term phage-bacteria coexistence, frequently observed in natural ecosystems (146), can result from CRISPR-mediated adaptive immunity (140) and that selective pressures imposed by these immune systems promote diversification of both phage and host populations (142).

Several metagenomic studies have shown that spacer sequences at the trailer end of the CRISPR are identical between strains of the same species over long time periods (141, 147, 148). This phenomenon, termed trailer-end clonality, seemed contradictory to the observed rapid pace of new spacer acquisition (131). Weinberger et al. (141) showed by mathematical modeling of virus and host populations that rapid selective sweeps of strains with successful CRISPR immunity against phages cause periodic elimination in trailer-end diversity. The preservation of trailer-end spacers over long time periods was predicted to protect the host against persisting, old viruses that occasionally rebloom.

The Influence of CRISPR on Phage Diversity

Genomes of viruses, and specifically phages, represent the most diverse sequence space on earth (149). The diversity of viruses may stem from the continuous selective pressure to adapt to bacterial resistance (8). For example, a common bacterial strategy for escaping viral attack involves mutations in the phage receptor (146) that force phages to diversify tail-fiber proteins that recognize cellular receptors (150). Although such receptor-driven selection can explain the huge diversity in phage tail-fiber sequences, the reason for the extreme rates of evolution across the entirety of the phage genome was obscure until the discovery of the CRISPR/Cas systems. Apart from the short PAM sequence, the CRISPR is largely indifferent to the gene or genomic position from which the spacer is taken. As a result, a strong selective pressure is continuously imposed on the genomes of invading parasites in an almost uniform distribution across the phage genome. Indeed, the first reports that CRISPR spacers protect against phages also noted that simple mutations and deletions in protospacers and PAMs gave rise to variant phages that remained infectious (80, 81). Similarly, phages exposed to continuous CRISPR surveillance show extensive patterns of recombination and shuffling of sequence motifs, presumably as a means to escape CRISPR resistance (131). Although other forces are also at work, this broad selective pressure is clearly an important contributor to the huge diversity among phages. On the flip side, phages are also primary mediators of host diversity, and DNA from phages and other genetic parasites often harbor genes with selectively advantageous traits. Gudbergsdottir et al. (97) have examined this genetic conflict by challenging cells with viruses or plasmids that carry a gene essential for cell growth. The cells that survive this challenge carry mutations that prevent crRNA-guided elimination of the beneficial DNA (e.g., deletion of spacers or shutdown of CRISPR transcription). Similarly, temperate phages may evade DNA-targeting CRISPR surveillance systems by integrating into the host chromosome (151). In this manner, phages may be major mediators of host evolution.

How much do we know about the phage sequence space worldwide? Early studies that inspected homologies between CRISPR spacers in microbial and phage genomes found matches to only 2% of all spacers, suggesting that most of the phage and plasmid sequences were still unexplored (76). However, as discussed above, we can now use CRISPR spacers as a tool to identify phage genomes in metagenomic analyses (131, 135). As a result, in metagenomic studies in which microbial communities are deeply sampled, much higher fractions of spacers match to known or predicted phage and plasmid sequences. For example, 35% of all spacers found in metagenomic sampling of the human gut microbiota had significant homology to contigs predicted to encode phage

and plasmid DNA (135). Therefore, continuous sampling of the CRISPR content in bacteria and archaea might significantly advance our appreciation of the sequence diversity of phages in natural microbial communities. Moreover, analysis of CRISPR spacers can provide information on the host range of specific phages.

The Role of Horizontal Gene Transfer in CRISPR-Phage Interactions

HGT enables sharing of DNA among species occupying the same niche. Phylogenetic studies of different *cas* genes strongly suggest that CRISPR/Cas systems have a high tendency for HGT (28, 29). Consistently, metagenomic studies document instances of CRISPR loci transfer (143). Because bacteria and archaea frequently exchange genetic information through large-scale recombination events (152, 153), recombination-based transfer of CRISPR arrays among strains in a community may allow sharing of successful immune repertoires. Indeed, plasmids and other mobile elements can carry CRISPR arrays (154). Curiously, CRISPRs have been reported in phage genomes (155, 156), suggesting that CRISPR-carrying phages may introduce the CRISPR into infected bacteria as a means of competing with other phages (156). It is likely that CRISPR-carrying phages also carry genes that enable them to escape CRISPR resistance.

Phages themselves are agents of HGT. We have known for many years that some phages can package random pieces of the infected bacterial genome into their particles, promoting genetic exchange among infected bacteria (157). This phenomenon is widespread in phages. Although the evolutionary incentive for phages to carry random pieces of bacterial genomes is obscure, one may speculate that phages use this mechanism to counteract CRISPR activity. Specifically, acquisition of a spacer of bacterial origin in a previous round of infection will lead to self-targeting and may result in eventual CRISPR loss, a process known as CRISPR-mediated autoimmunity (96).

SUMMARY POINTS

1. Bacteria and archaea have evolved nucleic acid–based adaptive defense systems that regulate the exchange of foreign DNA.
2. The CRISPR/Cas systems are phylogenetically and functionally diverse, but each of these systems relies on three common steps: new sequence integration, CRISPR RNA biogenesis, and crRNA-guided target interference.
3. Viral predation has a profound impact on the composition and the behavior of microbial communities in every ecological setting, and CRISPR-mediated adaptive immune systems play a major role in regulating the dynamic equilibrium between bacterial populations and their parasites.
4. CRISPR-associated surveillance complexes are easily programmable molecular sleds that can target any sequence of choice. These complexes offer new opportunities for implementation in biotechnology.

FUTURE ISSUES

1. PAM sequences are critical for DNA recognition by some CRISPR/Cas systems, but the molecular patterns that trigger CRISPR/Cas expression remain largely unknown.

2. Notwithstanding recent advances in understanding the requirements for new sequence integration, we still know little about the molecular mechanisms associated with new spacer selection and the molecular mechanism of the priming phenomenon.
3. Finding complementary target sequences in a crowded intracellular environment is analogous to finding the proverbial needle in the haystack. We do not understand how crRNA-guided surveillance systems locate complementary target sequences with efficiencies that provide protection against rapidly replicating phages.
4. Some CRISPR/Cas systems target RNA substrates for cleavage, but do these systems also include a reverse transcriptase for the integration of spacers derived from RNA-based phages?
5. Viruses employ diverse strategies to escape immune system detection. Sequence mutations are one mechanism of escape, but are there other virally encoded immune system subversion strategies?

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank many in the CRISPR research community, especially MaryClare F. Rollins, Samuel H. Sternberg, Dipali G. Sashital, and Ryan N. Jackson for thoughtful feedback on this manuscript. Research in the Wiedenheft lab is supported by the National Institutes of Health-NCRR COBRE (GM 103500). C.M.L is supported by grants from the National Science Foundation (MCB-0628732 and MCB-0920312), and R.S. is supported by the European Research Council Starting Grant program (grant 260432), the Israeli Science Foundation (grant ISF-1303/12), the Leona M. and Harry B. Helmsley Charitable Trust, and by a Deutsch-Israelische Projektkooperation grant from the Deutsche Forschungsgemeinschaft.

LITERATURE CITED

1. Hankin EH. 1896. L'action bactericide des eaux de la Jumna et du Gange sur le vibrion du cholera. *Ann. Inst. Pasteur* 10:511–23
2. Bergh O, Borsheim KY, Bratbak G, Heldal M. 1989. High abundance of viruses found in aquatic environments. *Nature* 340:467–68
3. Suttle CA. 2005. Viruses in the sea. *Nature* 437:356–61
4. Weinbauer MG. 2004. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* 28:127–81
5. Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasić L, Thingstad TF, et al. 2009. Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* 7:828–36
6. Bikard D, Marraffini LA. 2011. Innate and adaptive immunity in bacteria: mechanisms of programmed genetic variation to fight bacteriophages. *Curr. Opin. Immunol.* 24:15–20
7. Deveau H, Garneau JE, Moineau S. 2010. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.* 64:475–93
8. Stern A, Sorek R. 2010. The phage-host arms race: shaping the evolution of microbes. *BioEssays* 33:43–51
9. Westra ER, Swarts D, Staals R, Jore MM, Brouns SJJ, van der Oost J. 2012. The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu. Rev. Genet.* 46:311–38

10. Horvath P, Barrangou R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327:167–70
11. Marraffini LA, Sontheimer EJ. 2010. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* 11:181–90
12. Pougach KS, Lopatina AV, Severinov KV. 2012. CRISPR adaptive immunity systems of prokaryotes. *Mol. Biol.* 46:175–82
13. Wiedenheft B, Sternberg SH, Doudna JA. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482:331–38
14. Kunin V, Sorek R, Hugenholtz P. 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* 8:R61
15. Godde JS, Bickerton A. 2006. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J. Mol. Evol.* 62:718–29
16. 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.* 36:244–46
17. 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.* 169:5429–33
18. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35:W52–57
19. Rousseau C, Gonnet M, Le Romancer M, Nicolas J. 2009. CRISPI: a CRISPR interactive database. *Bioinformatics* 25:3317–18
20. Palmer KL, Gilmore MS. 2010. Multidrug-resistant enterococci lack CRISPR-*cas*. *mBio* 1:e00227
21. Jansen R, Embden JD, Gaastra W, Schouls LM. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* 43:1565–75
22. Pourcel C, Salvignol G, Vergnaud G. 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151:653–63
23. Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, et al. 2012. Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol. Cell* 45:292–302
24. Lillestøl RK, Shah SA, Brügger K, Redder P, Phan H, et al. 2009. CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Mol. Microbiol.* 72:259–72
25. Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, et al. 2010. Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol. Microbiol.* 77:1367–79
26. Pul U, Wurm R, Arslan Z, Geissen R, Hofmann N, Wagner R. 2010. Identification and characterization of *E. coli* CRISPR-*cas* promoters and their silencing by H-NS. *Mol. Microbiol.* 75:1495–512
27. Yosef I, Goren MG, Qimron U. 2012. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* 40:5569–76
28. Haft DH, Selengut J, Mongodin EF, Nelson KE. 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* 1:e60
29. 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
30. Makarova KS, Wolf YI, van der Oost J, Koonin EV. 2009. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol. Direct* 4:29
31. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, et al. 2011. Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9:467–77
32. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, et al. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321:960–64
33. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, et al. 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* 18:529–36

34. Gesner EM, Schellenberg MJ, Garside EL, George MM, MacMillan AM. 2011. Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat. Struct. Mol. Biol.* 18:688–92
35. Sashital DG, Jinek M, Doudna JA. 2011. An RNA induced conformational change required for CRISPR RNA cleavage by the endonuclease Cse3. *Nat. Struct. Mol. Biol.* 18:680–87
36. Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJJ, et al. 2011. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* 477:486–89
37. Lintner NG, Kerou M, Brumfield SK, Graham S, Liu H, et al. 2011. Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J. Biol. Chem.* 286:21643–56
38. Chen Z, Yang H, Pavletich NP. 2008. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* 453:489–84
39. Nam KH, Haitjema C, Liu X, Ding F, Wang H, et al. 2012. Cas5d protein processes pre-crRNA and assembles into a Cascade-like interference complex in Subtype I-C/Dvulg CRISPR-Cas system. *Structure* 20:1574–84
40. Wiedenheft B, van Duijn E, Bultema J, Waghmare S, Zhou K, et al. 2011. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. USA* 108:10092–97
41. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, et al. 2011. CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* 471:602–7
42. Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, et al. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468:67–71
43. 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* 337:816–21
44. Marraffini LA, Sontheimer EJ. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322:1843–45
45. Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, et al. 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139:945–56
46. Zhang J, Rouillon C, Kerou M, Reeks J, Brugger K, et al. 2012. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol. Cell* 45:303–13
47. Zhang J, Kasciukovic T, White MF. 2012. The CRISPR associated protein Cas4 is a 5′ to 3′ DNA exonuclease with an iron-sulfur cluster. *PLoS ONE* 7:e47232
48. Agari Y, Sakamoto K, Tamakoshi M, Oshima T, Kuramitsu S, Shinkai A. 2010. Transcription profile of *Thermus thermophilus* CRISPR systems after phage infection. *J. Mol. Biol.* 395:270–81
49. Young JC, Dill BD, Pan C, Hettich RL, Banfield JF, et al. 2012. Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in *Streptococcus thermophilus*. *PLoS ONE* 7:e38077
50. Shinkai A, Kira S, Nakagawa N, Kashihara A, Kuramitsu S, Yokoyama S. 2007. Transcription activation mediated by a cyclic AMP receptor protein from *Thermus thermophilus* HB8. *J. Bacteriol.* 189:3891–901
51. Cocozaki AI, Ramia NF, Shao Y, Hale CR, Terns RM, et al. 2012. Structure of the Cmr2 subunit of the CRISPR-Cas RNA silencing complex. *Structure* 20:545–53
52. Zhu X, Ye K. 2012. Crystal structure of Cmr2 suggests a nucleotide cyclase-related enzyme in type III CRISPR-Cas systems. *FEBS Lett.* 586:939–45
53. Lintner NG, Frankel KA, Tsutakawa SE, Alsbury DL, Copie V, et al. 2011. The structure of the CRISPR-associated protein Csa3 provides insight into the regulation of the CRISPR/Cas system. *J. Mol. Biol.* 405:939–55
54. Medina-Aparicio L, Rebollar-Flores JE, Gallego-Hernández AL, Vázquez A, Olvera L, et al. 2011. The CRISPR/Cas immune system is an operon regulated by LeuO, H-NS, and leucine-responsive regulatory protein in *Salmonella enterica* serovar Typhi. *J. Bacteriol.* 193:2396–407
55. Westra ER, Pul U, Heidrich N, Jore MM, Lundgren M, et al. 2010. H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.* 77:1380–93
56. Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H, et al. 2006. Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science* 313:236–38

57. Navarre WW, McClelland M, Libby SJ, Fang FC. 2007. Silencing of xenogeneic DNA by H-NS-facilitation of lateral gene transfer in bacteria by a defense system that recognizes foreign DNA. *Genes Dev.* 21:1456–71
58. Bouffartigues E, Buckle M, Badaut C, Travers A, Rimsky S. 2007. H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing. *Nat. Struct. Mol. Biol.* 14:441–48
59. Lang B, Blot N, Bouffartigues E, Buckle M, Geertz M, et al. 2007. High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. *Nucleic Acids Res.* 35:6330–37
60. Liu YJ, Chen H, Kenney LJ, Yan J. 2010. A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes Dev.* 24:339–44
61. Hernández-Lucas I, Gallego-Hernández AL, Encarnación S, Fernández-Mora M, Martínez-Batallar AG, et al. 2008. The LysR-type transcriptional regulator LeuO controls expression of several genes in *Salmonella enterica* serovar Typhi. *J. Bacteriol.* 190:1658–70
62. Fang M, Majumder A, Tsai KJ, Wu HY. 2000. ppGpp-dependent *leuO* expression in bacteria under stress. *Biochem. Biophys. Res. Commun.* 276:64–70
63. Majumder A, Fang M, Tsai KJ, Ueguchi C, Mizuno T, Wu HY. 2001. LeuO expression in response to starvation for branched-chain amino acids. *J. Biol. Chem.* 276:19046–51
64. Potrykus K, Cashel M. 2008. (p)ppGpp: still magical? *Annu. Rev. Microbiol.* 62:35–51
65. Tosa T, Pizer LI. 1971. Biochemical bases for the antimetabolite action of L-serine hydroxamate. *J. Bacteriol.* 106:972–82
66. Williams E, Lowe TM, Savas J, DiRuggiero J. 2007. Microarray analysis of the hyperthermophilic archaeon *Pyrococcus furiosus* exposed to gamma irradiation. *Extremophiles* 11:19–29
67. Strand KR, Sun C, Li T, Jenney FE Jr, Schut GJ, Adams MW. 2010. Oxidative stress protection and the repair response to hydrogen peroxide in the hyperthermophilic archaeon *Pyrococcus furiosus* and in related species. *Arch. Microbiol.* 192:447–59
68. Plagens A, Tjaden B, Hagemann A, Randau L, Hensel R. 2012. Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. *J. Bacteriol.* 194:2491–500
69. Götz D, Paytubi S, Munro S, Lundgren M, Bernander R, White MF. 2007. Responses of hyperthermophilic crenarchaea to UV irradiation. *Genome Biol.* 8:R220
70. Perez-Rodriguez R, Haitjema C, Huang Q, Nam KH, Bernardis S, et al. 2011. Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in *Escherichia coli*. *Mol. Microbiol.* 79:584–99
71. Juranek S, Eban T, Altuvia Y, Brown M, Morozov P, et al. 2012. A genome-wide view of the expression and processing patterns of *Thermus thermophilus* HB8 CRISPR RNAs. *RNA* 18:783–94
72. Tang TH, Polacek N, Zywicki M, Huber H, Brugger K, et al. 2005. Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* 55:469–81
73. Tang TH, Bachellerie JP, Rozhdestvensky T, Bortolin ML, Huber H, et al. 2002. Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl. Acad. Sci. USA* 99:7536–41
74. Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R. 2010. A single-base resolution map of an archaeal transcriptome. *Genome Res.* 20:133–41
75. Bolotin A, Ouinquis B, Sorokin A, Ehrlich SD. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151:2551–61
76. Mojica FJ, Díez-Villaseñor C, García-Martínez J, Soria E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* 60:174–82
77. Barrangou R, Horvath P. 2012. CRISPR: new horizons in phage resistance and strain identification. *Annu. Rev. Food Sci. Technol.* 3:143–62
78. Brussow H. 2001. Phages of dairy bacteria. *Annu. Rev. Microbiol.* 55:283–303
79. Moineau S, Lévesque C. 2005. Control of bacteriophages in industrial fermentations. *Bacteriophages: Biology and Applications*, ed. E Kutter, A Sulakvelidze, pp. 282–93. Boca Raton, FL: CRC
80. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315:1709–12

81. Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, et al. 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* 190:1390–400
82. Horvath P, Romero DA, Couté-Monvoisin AC, Richards M, Deveau H, et al. 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* 190:1401–12
83. Mojica FJ, Díez-Villaseñor C, García-Martínez J, Almendros C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155:733–40
84. Almendros C, Guzmán NM, Díez-Villaseñor C, García-Martínez J, Mojica FJ. 2012. Target motifs affecting natural immunity by a constitutive CRISPR-Cas system in *Escherichia coli*. *PLoS ONE* 7:e50797
85. Cady KC, Bondy-Denomy J, Heussler GE, Davidson AR, O'Toole GA. 2012. The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. *J. Bacteriol.* 194:5728–38
86. van der Ploeg JR. 2009. Analysis of CRISPR in *Streptococcus mutans* suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. *Microbiology* 155:1966–76
87. Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* 39:9275–82
88. Ellinger P, Arslan Z, Wurm R, Tschapek B, Mackenzie C, et al. 2012. The crystal structure of the CRISPR-associated protein Csn2 from *Streptococcus agalactiae*. *J. Struct. Biol.* 178:350–62
89. Koo Y, Jung DK, Bae E. 2012. Crystal structure of *Streptococcus pyogenes* Csn2 reveals calcium-dependent conformational changes in its tertiary and quaternary structure. *PLoS ONE* 7:e33401
90. Nam KH, Kurinov I, Ke AL. 2011. Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca^{2+}-dependent double-stranded DNA binding activity. *J. Biol. Chem.* 286:30759–68
91. 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
92. Swarts DC, Mosterd C, van Passel MW, Brouns SJ. 2012. CRISPR interference directs strand specific spacer acquisition. *PLoS ONE* 7:e35888
93. Babu M, Beloglazova N, Flick R, Graham C, Skarina T, et al. 2011. A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Mol. Microbiol.* 79:484–502
94. Beloglazova N, Brown G, Zimmerman MD, Proudfoot M, Makarova KS, et al. 2008. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol. Chem.* 283:20361–71
95. Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W, Doudna JA. 2009. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated antiviral defense. *Structure* 17:904–12
96. Stern A, Keren L, Wurtzel O, Amitai G, Sorek R. 2010. Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet.* 26:335–40
97. Gudbergsdottir S, Deng L, Chen Z, Jensen JV, Jensen LR, et al. 2011. Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* 79:35–49
98. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, et al. 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. USA* 108:10098–103
99. Carte J, Pfister NT, Compton MM, Terns RM, Terns MP. 2010. Binding and cleavage of CRISPR RNA by Cas6. *RNA* 16:2181–88
100. Carte J, Wang R, Li H, Terns RM, Terns MP. 2008. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* 22:3489–96
101. Hatoum-Aslan A, Maniv I, Marraffini LA. 2011. Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc. Natl. Acad. Sci. USA* 108:21218–22
102. Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA. 2010. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329:1355–58
103. Haurwitz RE, Sternberg SH, Doudna JA. 2012. Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *EMBO J.* 31:2824–32

104. Sternberg SH, Haurwitz RE, Doudna JA. 2012. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* 18:661–72
105. Wang RY, Li H. 2012. The mysterious RAMP proteins and their roles in small RNA-based immunity. *Protein Sci.* 21:463–70
106. Maris C, Dominguez C, Allain FHT. 2005. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* 272:2118–31
107. Ebihara A, Yao M, Masui R, Tanaka I, Yokoyama S, Kuramitsu S. 2006. Crystal structure of hypothetical protein TTHB192 from *Thermus thermophilus* HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein Sci.* 15:1494–99
108. Wang R, Preamplume G, Terns MP, Terns RM, Li H. 2011. Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure* 19:257–64
109. Wang R, Zheng H, Preamplume G, Shao Y, Li H. 2012. The impact of CRISPR repeat sequence on structures of a Cas6 protein-RNA complex. *Protein Sci.* 21:405–17
110. Hale C, Kleppe K, Terns RM, Terns MP. 2008. Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* 14:2572–79
111. Calnan BJ, Tidor B, Biancalana S, Hudson D, Frankel AD. 1991. Arginine-mediated RNA recognition: the arginine fork. *Science* 252:1167–71
112. Yang W. 2011. Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.* 44:1–93
113. Halford SE, Marko JF. 2004. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* 32:3040–52
114. Westra ER, van Erp PBG, Künne T, Wong SP, Staals RHJ, et al. 2012. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell* 46:595–605
115. Gowers DM, Halford SE. 2003. Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *EMBO J.* 22:1410–18
116. Sashital DG, Wiedenheft B, Doudna JA. 2012. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol. Cell* 48:606–15
117. Mulepati S, Orr A, Bailey S. 2012. Crystal structure of the largest subunit of a bacterial RNA-guided immune complex and its role in DNA target binding. *J. Biol. Chem.* 287:22445–49
118. Banerjee A, Santos WL, Verdine GL. 2006. Structure of a DNA glycosylase searching for lesions. *Science* 311:1153–57
119. Horton JR, Zhang X, Maunus R, Yang Z, Wilson GG, et al. 2006. DNA nicking by HinP1I endonuclease: bending, base flipping and minor groove expansion. *Nucleic Acids Res.* 34:939–48
120. Westra ER, Nilges B, van Erp PBG, van der Oost J, Dame RT, Brouns SJJ. 2012. Cascade-mediated binding and bending of negatively supercoiled DNA. *RNA Biol.* 9:1134–1138
121. Beloglazova N, Petit P, Flick R, Brown G, Savchenko A, Yakunin AF. 2011. Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO J.* 30:4616–27
122. Han D, Krauss G. 2009. Characterization of the endonuclease SSO2001 from *Sulfolobus solfataricus* P2. *FEBS Lett.* 583:771–76
123. Mulepati S, Bailey S. 2011. Structural and biochemical analysis of the nuclease domain of the clustered regularly interspaced short palindromic repeat (CRISPR) associated protein 3 (CAS3). *J. Biol. Chem.* 286:31896–903
124. Sinkunas T, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. 2011. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* 30:1335–42
125. Marraffini LA, Sontheimer EJ. 2010. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 463:568–71
126. Sakamoto K, Agari Y, Agari K, Yokoyama S, Kuramitsu S, Shinkai A. 2009. X-ray crystal structure of a CRISPR-associated RAMP module Cmr5 protein from *Thermus thermophilus* HB8. *Proteins* 75:528–32
127. Bolduc B, Shaughnessy DP, Wolf YI, Koonin EV, Roberto FF, Young M. 2012. Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated Yellowstone hot springs. *J. Virol.* 86:5562–73
128. Dennehy JJ. 2009. Bacteriophages as model organisms for virus emergence research. *Trends Microbiol.* 17:450–57

129. Gómez P, Buckling A. 2011. Bacteria-phage antagonistic coevolution in soil. *Science* 332:106–9
130. Vos M, Birkett PJ, Birch E, Griffiths RI, Buckling A. 2009. Local adaptation of bacteriophages to their bacterial hosts in soil. *Science* 325:833
131. Andersson AF, Banfield JF. 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 320:1047–50
132. Heidelberg JF, Nelson WC, Schoenfeld T, Bhaya D. 2009. Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes. *PLoS ONE* 4:e4169
133. Held NL, Whitaker RJ. 2009. Viral biogeography revealed by signatures in *Sulfolobus islandicus* genomes. *Environ. Microbiol.* 11:457–66
134. Pride DT, Sun CL, Salzman J, Rao N, Loomer P, et al. 2012. Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. *Genome Res.* 21:126–36
135. Stern A, Mick E, Tirosh I, Sagy O, Sorek R. 2012. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Res.* 10:1985–94
136. Rho M, Wu YW, Tang H, Doak TG, Ye Y. 2012. Diverse CRISPRs evolving in human microbiomes. *PLoS Genet.* 8:e1002441
137. Sorokin VA, Gelfand MS, Artamonova II. 2010. Evolutionary dynamics of clustered irregularly interspaced short palindromic repeat systems in the ocean metagenome. *Appl. Environ. Microbiol.* 76:2136–44
138. Levin BR. 2010. Nasty viruses, costly plasmids, population dynamics, and the conditions for establishing and maintaining CRISPR-mediated adaptive immunity in bacteria. *PLoS Genet.* 6:e1001171
139. He J, Deem MW. 2010. Heterogeneous diversity of spacers within CRISPR (clustered regularly interspaced short palindromic repeats). *Phys. Rev. Lett.* 105:128102
140. Haerter JO, Trusina A, Sneppen K. 2011. Targeted bacterial immunity buffers phage diversity. *J. Virol.* 85:10554–60
141. Weinberger AD, Sun CL, Plucinski MM, Denef VJ, Thomas BC, et al. 2012. Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Comput. Biol.* 8:e1002475
142. Childs LM, Held NL, Young MJ, Whitaker RJ, Weitz JS. 2012. Multiscale model of CRISPR-induced coevolutionary dynamics: diversification at the interface of Lamarck and Darwin. *Evolution* 66:2015–29
143. Tyson GW, Banfield JF. 2008. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* 10:200–7
144. Tringe SG, Rubin EM. 2005. Metagenomics: DNA sequencing of environmental samples. *Nat. Rev. Genet.* 6:805–14
145. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59–65
146. Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D. 2011. Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* 474:604–8
147. Díez-Villaseñor C, Almendros C, García-Martínez J, Mojica FJ. 2010. Diversity of CRISPR loci in *Escherichia coli*. *Microbiology* 156:1351–61
148. Held NL, Herrera A, Cadillo-Quiroz H, Whitaker RJ. 2010. CRISPR associated diversity within a population of *Sulfolobus islandicus*. *PLoS ONE* 5:e12988
149. Hambly E, Suttle CA. 2005. The viriosphere, diversity, and genetic exchange within phage communities. *Curr. Opin. Microbiol.* 8:444–50
150. Medhekar B, Miller JF. 2007. Diversity-generating retroelements. *Curr. Opin. Microbiol.* 10:388–95
151. Dyall-Smith M. 2011. Dangerous weapons: a cautionary tale of CRISPR defence. *Mol. Microbiol.* 79:3–6
152. Denef VJ, Kalnejais LH, Mueller RS, Wilmes P, Baker BJ, et al. 2010. Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. *Proc. Natl. Acad. Sci. USA* 107:2383–90
153. 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.* 22:1444–48
154. Guo P, Cheng Q, Xie P, Fan Y, Jiang W, Qin Z. 2011. Characterization of the multiple CRISPR loci on *Streptomyces* linear plasmid pSHK1. *Acta Biochim. Biophys. Sin.* 43:630–39

155. Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, et al. 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.* 38:779–86
156. Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, et al. 2011. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* 21:1616–25
157. Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brussow H. 2003. Phage as agents of lateral gene transfer. *Curr. Opin. Microbiol.* 6:417–24
158. Samai P, Smith P, Shuman S. 2010. Structure of a CRISPR-associated protein Cas2 from *Desulfovibrio vulgaris*. *Acta Crystallogr. Sect. F* 66:1552–56