REVIEW

doi:10.1038/nature10886

RNA-guided genetic silencing systems in bacteria and archaea

Blake Wiedenheft[1,2]†, Samuel H. Sternberg[3] & Jennifer A. Doudna[1–4]

Clustered regularly interspaced short palindromic repeat (CRISPR) are essential components of nucleic-acid-based adaptive immune systems that are widespread in bacteria and archaea. Similar to RNA interference (RNAi) pathways in eukaryotes, CRISPR-mediated immune systems rely on small RNAs for sequence-specific detection and silencing of foreign nucleic acids, including viruses and plasmids. However, the mechanism of RNA-based bacterial immunity is distinct from RNAi. Understanding how small RNAs are used to find and destroy foreign nucleic acids will provide new insights into the diverse mechanisms of RNA-controlled genetic silencing systems.

Bacteria and archaea are the most diverse and abundant organisms on the planet, thriving in habitats that range from hot springs to humans. However, viruses outnumber their microbial hosts in every ecological setting, and the selective pressures imposed by these rapidly evolving parasites has driven the diversification of microbial defence systems[1–3]. Historically, our understanding of antiviral immunity in bacteria has focused on restriction-modification systems, abortive-phage phenotypes, toxin–antitoxins and other innate defence systems[4,5]. More recently, bioinformatic, genetic and biochemical studies have revealed that many prokaryotes use an RNA-based adaptive immune system to target and destroy genetic parasites (reviewed in refs 6–12). Such adaptive immunity, previously thought to occur only in eukaryotes, provides an example of RNA-guided destruction of foreign genetic material by a process that is distinct from RNA interference (RNAi) (Fig. 1).

In response to viral and plasmid challenges, bacteria and archaea integrate short fragments of foreign nucleic acid into the host chromosome at one end of a repetitive element known as CRISPR (clustered regularly interspaced short palindromic repeat)[13–15]. These repetitive loci serve as molecular 'vaccination cards' by maintaining a genetic record of prior encounters with foreign transgressors. CRISPR loci are transcribed, and the long primary transcript is processed into a library of short CRISPR-derived RNAs (crRNAs)[16–21] that each contain a sequence complementary to a previously encountered invading nucleic acid. Each crRNAs is packaged into a large surveillance complex that patrols the intracellular environment and mediates the detection and destruction of foreign nucleic acid targets[15,22–27].

CRISPRs were originally identified in the *Escherichia coli* genome in 1987, when they were described as an unusual sequence element consisting of a series of 29-nucleotide repeats separated by unique 32-nucleotide 'spacer' sequences[28]. Repetitive sequences with a similar repeat–spacer–repeat pattern were later identified in phylogenetically diverse bacterial and archaeal genomes, but the function of these repeats remained obscure until many spacer sequences were recoginized as being identical to viral and plasmid sequences[29–31]. This observation led to the hypothesis that CRISPRs provide a genetic memory of infection[29], and the detection of short CRISPR-derived RNA transcripts suggested that there may be functional similarities between CRISPR-based immunity and RNAi[30,32]. In this Insight, we review three stages of CRISPR-based adaptive immunity and compare mechanistic aspects of these immune systems to other RNA-guided genetic silencing pathways.

Architecture and composition of CRISPR loci

The defining feature of CRISPR loci is a series of direct repeats (approximately 20–50 base pairs) separated by unique spacer sequences of a similar length[11,33,34] (Fig. 2). The repeat sequences within a CRISPR locus are conserved, but repeats in different CRISPR loci can vary in both sequence and length. In addition, the number of repeat–spacer units in a CRISPR locus varies widely within and among organisms[35].

The sequence diversity of these repetitive loci initially limited their detection and obscured their relationship, but computational methods have been developed for detecting repeat patterns rather than related sequences[33,34,36–38]. One of the first-generation pattern-recognition algorithms identified the repeat–spacer–repeat architecture in phylogenetically diverse bacterial and archaeal genomes, but related structures were not identified in eukaryotic chromosomes[39]. Comparative analyses of the sequences adjacent to the CRISPR loci have revealed an (A+T)-rich 'leader' sequence that has been shown to serve as a promoter element for CRISPR transcription[39–42]. In addition to the leader sequence, Jansen *et al.*[39] identified a set of four CRISPR-associated (*cas*) genes known as *cas*1–4 that are found exclusively in genomes containing CRISPRs. Based on sequence similarity to proteins of known function, Cas3 was predicted to be a helicase and Cas4 a RecB-like exonuclease[39].

Subsequent bioinformatic analyses have shown that CRISPR loci are flanked by a large number of extremely diverse *cas* genes[32,43]. The *cas1* gene is a common component of all CRISPR systems, and phylogenetic analyses of Cas1 sequences indicate there are several versions of the CRISPR system. Providing additional evidence for the classification of distinct CRISPR types, neighbourhood analysis has identified conserved arrangements of between four and ten *cas* genes that are found in association with CRISPR loci harbouring specific repeat sequences[35].

These distinct immune systems have been divided into three major CRISPR types on the basis of gene conservation and locus organization[10]. More than one CRISPR type is often found in a single organism, indicating that these systems are probably mutually compatible and could share functional components[10]. Despite the variation in number and diversity of *cas* genes, the distinguishing feature of all type I systems is that they encode a *cas3* gene. The Cas3 protein contains an N-terminal HD phosphohydrolase domain and a C-terminal helicase domain[32,39,43,44]. In some type I systems, the Cas3 nuclease and helicase domains are encoded by separate genes (*cas3″* and *cas3′*, respectively), but in each case they are thought to participate in degrading foreign nucleic acids[22,44–46] (Fig. 2).

[1]Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, Maryland 20815-6789, USA. [2]Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. [3]Department of Chemistry, University of California, Berkeley, California 94720, USA. [4]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.
†Present address: Department of Immunology and Infectious Diseases, Montana State University, Bozeman, Montana, USA

Type II CRISPR systems consist of just four *cas* genes, one of which is always *cas9* (formerly referred to as *csn1*). Cas9 is a large protein that includes both a RuvC-like nuclease domain and an HNH nuclease domain. Studies in *Streptococcus pyogenes* and *Streptococcus thermophilus* have indicated that Cas9 may participate in both CRISPR RNA processing and target destruction[14,15,17]. Two variations of the type III system have been identified (known as III-A and III-B). This division is supported by the functional differences reported in *Staphylococcus epidermidis* and *Pyrococcus furiosus*[47,48]. The immune system in *S. epidermidis* (type III-A) targets plasmid DNA *in vivo*, whereas the purified components of the type III-B system in *P. furiosus* have been found to cleave only single-stranded RNA substrates *in vitro*. The functional distinction between these two closely related systems suggests there could be other mechanistic differences between the distinct CRISPR subtypes.

Integration of new information into CRISPR loci

Acquisition of foreign DNA is the first step of CRISPR-mediated immunity (Fig. 2 and 3). During this stage, a short segment of DNA from an invading virus or plasmid (known as the protospacer) is integrated preferentially at the leader end of the CRISPR locus[14,15]. Although metagenomic studies performed on environmental samples indicate that CRISPRs evolve rapidly in dynamic equilibrium with resident phage populations[13,49,50], the type II system in *S. thermophilus* is currently the only CRISPR system that has been shown to robustly acquire new phage or plasmid sequences in a pure culture. Phage-challenge experiments in *S. thermophilus* have indicated that a small proportion of the cells in a population will typically incorporate a single virus-derived sequence at the leader end of a CRISPR locus[14,15,51,52]. The CRISPR-repeat sequence is duplicated for each new spacer seqenced added, thus maintaining the repeat–spacer–repeat architecture. Although the mechanism of spacer integration and replication of the repeat sequence is still unknown, studies in *S. thermophilus* and *E. coli* have indicated that several Cas proteins are involved in the process[14,15,22,53]. Mutational analysis of the *cas* genes in *S. thermophilus* demonstrated that *csn2* (previously known as *cas7*) is required for new spacer sequence acquisition[14]. This gene is not conserved in other CRISPR types, which suggests that either the mechanism of adaptation in *S. thermophilus* is distinct from the other types or that there are functional orthologues of Csn2 in other systems. Furthermore, gene deletion experiments in both *S. thermophilus* and *E. coli* have shown that neither *cas1* nor *cas2* genes are required for CRISPR RNA processing or targeted interference[22,53,54]. These genetic studies suggest a role for Cas1 and Cas2 in the integration of foreign DNA into the CRISPR.

The role of Cas1 in CRISPR-mediated immunity is still uncertain; however, biochemical and structural data indicate a function for Cas1 in new–spacer–sequence acquisition[54–56]. Cas1 proteins from *Pseudomonas aeruginosa*[56], *E. coli*[54] and *Sulfolobus solfataricus*[55] have been purified and studied biochemically. The Cas1 protein from *S. solfataricus* has been shown to bind nucleic acids with high affinity (K_d ranging from 20 to 50 nM), but without sequence preference[55]. The Cas1 protein from *E. coli* also binds to DNA with a preference for mismatched or abasic substrates[57]. This observation is consistent with a recent study showing a physical and genetic interaction between *E. coli* Cas1 and several proteins associated with DNA replication and repair[54].

Activity assays with Cas1 from *P. aeruginosa* and *E. coli* indicate that Cas1 is a metal-dependent nuclease. The Cas1 protein from *P. aeruginosa* is a DNA-specific nuclease, whereas the Cas1 protein from *E. coli* had a nuclease activity on a wider range of nucleic acid substrates[54,56]. These *in vitro* assays suggest that Cas1 proteins interact with nucleic acids in a non-sequence-specific manner.

Crystal structures for five different Cas1 proteins are currently available (Protein Data Bank (PDB) identifiers: 3GOD, 3NKD, 3LFX, 3PV9 and 2YZS)[54,56]. Although the amino acid sequences for these proteins are extremely diverse (less than 15% sequence identity), their tertiary and quaternary structures are similar. All Cas1 proteins seem to share a two-domain architecture consisting of an N-terminal β-strand domain and a

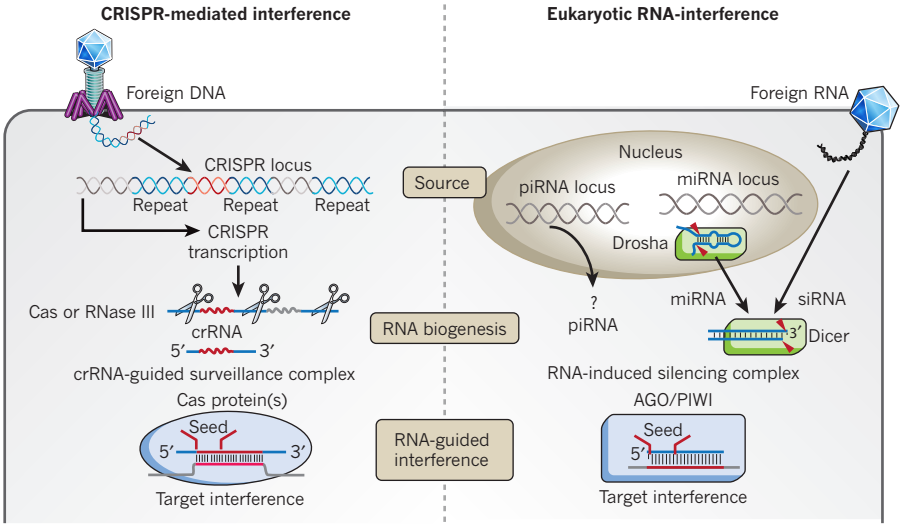

Figure 1 | Parallels and distinctions between CRISPR RNA-guided silencing systems and RNAi. CRISPR systems and RNAi recognize long RNA precursors that are processed into small RNAs, which act as sequence-specific guides for targeting complementary nucleic acids. In CRISPR systems, foreign DNA is integrated into the CRISPR locus, and long transcripts from these loci are processed by a CRISPR-associated (Cas) or RNase III family nuclease[16–21,64]. The short CRISPR-derived RNAs (crRNAs) assemble with Cas proteins into large surveillance complexes that target destruction of invading genetic material[15,22,24–27,48]. In some eukaryotes, long double-stranded RNAs are recognized as foreign, and a specialized RNase III family endoribonuclease (Dicer) cleaves these RNAs into short-interfering RNAs (siRNAs) that guide the immune system to invading RNA viruses[76]. PIWI-interacting RNAs (piRNAs) are transcribed from repetitive clusters in the genome that often contain many copies of retrotransposons and primarily act by restricting transposon mobility[76–78]. The biogenesis of piRNAs is not yet fully understood. MicroRNAs (miRNAs) are also encoded on the chromosome, and primary miRNA transcripts form stable hairpin structures that are sequentially processed (shown by red triangles) by two RNase III family endoribonucleases (Drosha and Dicer)[79]. miRNAs do not participate in genome defence but are major regulators of endogenous gene expression[80]. Like crRNAs, eukaryotic piRNAs, siRNAs and miRNAs associate with proteins that facilitate complementary interactions with invading nucleic acid targets[27,60,69,79]. In eukaryotes, the Argonaute proteins pre-order the 5′ region of the guide RNA into a helical configuration, reducing the entropy penalty of interactions with target RNAs[69]. This high-affinity binding site, called the 'seed' sequence, is essential for target sequence interactions. Recent studies indicate that the CRISPR system may use a similar seed-binding mechanism for enhancing target sequence interactions[26,27,53,60].

C-terminal α-helical domain (Fig. 3). The C-terminal domain contains a conserved divalent metal-ion binding site, and alanine substitutions of the metal-coordinating residues inhibit Cas1-catalysed DNA degradation[54,56]. The metal ion is surrounded by a cluster of basic residues that form a strip of positive charge across the surface of the C-terminal domain. This positively charged surface may serve as an electrostatic snare to position nucleic-acid substrates near the catalytic metal ions[56] (Fig. 3). The Cas1 protein forms a stable homodimer that is formed through interactions between the two β-strand domains, which are related by a pseudo-two-fold axis of symmetry[54,56]. This organization creates a saddle-like structure that can be modelled onto double-stranded DNA without steric clashing. β-hairpins, one from each of the two symmetrically related molecules, hang on opposite faces of the double-stranded DNA (like stirrups on a saddle). Although this feature of the Cas1 structure did not initially stand out as a potential DNA-binding site, comparative analysis of the available Cas1 structures reveals a conserved set of positively charged residues along each of the β-hairpins that could contact the phosphate backbone. The two β-hairpins, which are symmetrically related, might participate in sequence-specific interactions with the CRISPR repeat, whereas the large positively charged surface on the C-terminal α-helical domain could account for the high-affinity, non-sequence-specific interactions that have been observed *in vitro*.

In spite of these structural studies and biochemical results, it is still only possible to speculate on the role of Cas1 in the integration of new spacer sequences, and many steps associated with the integration process still need to be explained. For example, new spacer sequences are inserted

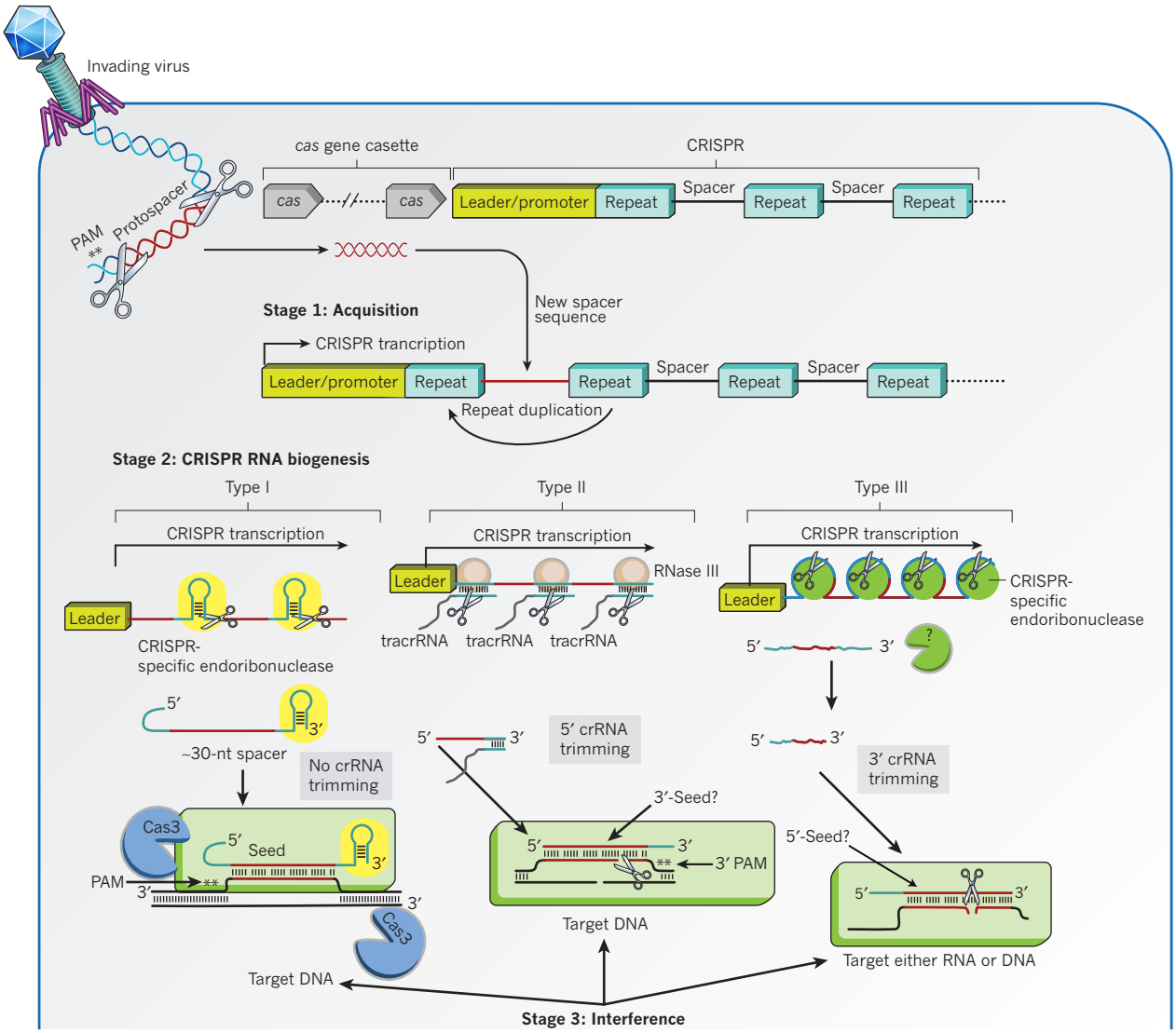

Figure 2 | Diversity of CRISPR-mediated adaptive immune systems in bacteria and archaea. A diverse set of CRISPR-associated (*cas*) genes (grey arrows) encode proteins required for new spacer sequence acquisition (Stage 1), CRISPR RNA biogenesis (Stage 2) and target interference (Stage 3). Each CRISPR locus consists of a series of direct repeats separated by unique spacer sequences acquired from invading genetic elements (protospacers). Protospacers are flanked by a short motif called the protospacer adjacent motif (PAM, **) that is located on the 5′ (type I) or 3′ (type II) side in foreign DNA[10,51,52,59,67]. Long CRISPR transcripts are processed into short crRNAs by distinct mechanisms. In type I and III systems, a CRISPR-specific endoribonuclease (yellow ovals and green circles, respectively) cleaves 8 nucleotides upstream of each spacer sequence[16,18–21,64]. In type III systems, the repeat sequence on the 3′ end of the crRNA is trimmed by an unknown mechanism (green pacman, right). In type II systems, a *trans*-acting antisense RNA (tracrRNA) with complementarity to the CRISPR RNA repeat sequence forms an RNA duplex that is recognized and cleaved by cellular RNase III (brown ovals)[17]. This cleavage intermediate is further processed at the 5′ end resulting in a mature, approximately 40-nucleotide crRNA with an approximately 20-nucleotide 3′-handle. In each system, the mature crRNA associates with one or more Cas proteins to form a surveillance complex (green rectangles). Type I systems encode a Cas3 nuclease (blue pacman), which may be recruited to the surveillance complex following target binding[24,27,44]. A short high-affinity binding site called a seed-sequence has been identified in some type I systems[27,60], and genetic experiments suggest that type II systems have a seed sequence located at the 3′ end of the crRNA spacer sequence[53].

preferentially at the leader end of the CRISPR, but the mechanism of leader end recognition is unknown. One simple model suggests that the leader sequence contains a recognition element that recruits the integration machinery. It is equally possible that integration relies on single-stranded regions of the CRISPR DNA that are made available during transcription. Transcription-associated recombination is involved in genome stability[58], and a mechanism that couples integration together with transcription would link the process of adaptation to CRISPR RNA expression, ensuring that spacers from the most recent virus or plasmid are transcribed first.

The integration machinery must be able to distinguish foreign DNA from that of the host genome. The molecular cues that are involved in the distinction of 'self' from 'non-self' are still unknown, but sequencing of CRISPR loci following phage challenge suggests that spacer sequences are not selected at random[15,29,51,52,59,60]. Mapping spacer sequences onto viral genomes reveals a short sequence motif proximal to the protospacer, which is referred to as the protospacer adjacent motif (PAM). PAM sequences are only a few nucleotides long, and the precise sequence varies depending on the CRISPR system type[59]. This variation suggests that one or more of the Cas proteins associated with each immune system is involved in PAM recognition, but the mechanism governing this specificity is unknown.

CRISPR RNA biogenesis

Spacer acquisition is the first step of immunization, but successful protection from bacteriophage or plasmid challenge requires the CRISPR to be transcribed and processed into short CRISPR-derived RNAs (crRNAs). crRNAs were first detected by small RNA profiling in *Archaeoglobus fulgidus*[61] and *S. solfataricus*[62]. Northern-blot analysis using probes against the repeat sequence of the CRISPR revealed a 'ladder-like' pattern of RNA consistent with a long precursor CRISPR RNA transcript (pre-crRNA) that was processed at approximately 60-nucleotide intervals. In fact, the 3′ ends of cloned crRNAs were mapped to the middle of the CRISPR repeat[61], which suggested that the repeat sequence was recognized and cleaved.

The need for crRNAs in CRISPR-mediated defence was demonstrated initially by investigation of a CRISPR-specific endoribonuclease in *E. coli* called Cas6e (formerly known as Cse3 or CasE)[22]. Cas6e specifically binds and cleaves within each repeat sequence of the long pre-crRNA, resulting in a library of crRNAs that each contain a unique spacer sequence flanked by fragments of the adjacent repeats. Mutation of a conserved histidine blocks crRNA biogenesis and leaves the cell susceptible to phage infection[22].

The Cas6e protein consists of a double ferredoxin-like fold that selectively associates with specific RNA repeats and does not associate with DNA or CRISPR RNAs containing a non-cognate repeat sequence [18,20,22,63] (Fig. 4). Crystal structures of Cas6e bound to a CRISPR RNA repeat reveal a combination of sequence- and structure-specific interactions that explain the molecular mechanism of substrate recognition[18,20]. The repeat sequence of the *E. coli* CRISPR is partially palindromic, and the RNA forms a stable (approximately 20-nucleotide) stem loop[22,35]. A positively charged β-hairpin in Cas6e interacts with the major groove of the RNA duplex, which positions the 3′ strand of the crRNA stem along a conserved, positively charged cleft on one face of the protein[18,20] (Fig. 4). RNA binding induces a conformational change that disrupts the bottom base pair of the stem and positions the scissile phosphate within the enzyme active site for site-specific cleavage[20]. CRISPR RNA cleavage occurs 8 nucleotides upstream of the spacer sequence, which results in 61-nucleotide mature crRNAs consisting of a 32-nucleotide spacer flanked by 8 nucleotides of the repeat sequence on the 5′ end (known as the 5′-handle) and 21 nucleotides of the remaining repeat sequence on the 3′ end (Fig. 4). Cas6e remains tightly bound to the 3′ stem-loop[20] and may serve as a nucleation point for assembly of a large effector complex, Cascade (CRISPR-associated complex for antiviral defence), that is required for phage silencing in the next stage of the immune system[22,24,26] (discussed later).

Crystal structures of CRISPR-specific endoribonucleases from two other immune systems offer additional insights into the co-evolutionary relationship between these specialized enzymes and their cognate RNAs[16,19,21] (Fig. 4). In *P. aeruginosa*, Cas6f (previously known as Csy4) specifically binds and cleaves the CRISPR-RNA-repeat 8 nucleotides upstream of the spacer sequence, which leaves a similar 8-nucleotide 5′-handle on mature crRNAs[19]. The co-crystal structure of Cas6f bound to its cognate RNA reveals interesting parallels between the method of RNA binding used by Cas6f and Cas6e[18–20]. Like Cas6e, the *P. aeruginosa* Cas6f protein recognizes the sequence and shape of a stable stem-loop in the crRNA repeat sequence by interacting extensively with the major groove of the double-stranded RNA. However, the structural elements responsible for this interaction are distinct between the two proteins[18–20] (Fig. 4). The Cas6f protein has a two-domain architecture, which consists of an N-terminal ferredoxin-like fold similar to that in Cas6e, but its C-terminal domain is structurally distinct. An arginine-rich helix in the C-terminal domain of Cas6f inserts into the major groove of the crRNA duplex, and the bottom of the crRNA is positioned for sequence-specific hydrogen-bonding contacts in the RNA major groove. These contacts position the scissile phosphate of the crRNA in the enzyme active site so that cleavage occurs 8 nucleotides upstream of the spacer sequence[18–20] (Fig. 4).

Although Cas6f and Cas6e recognize the sequence and shape of the crRNA hairpin in their respective systems, CRISPR RNA repeats in other CRISPR systems are thought to be unstructured[35]. For example, the Cas6 protein from *P. furiosus* associates with CRISPR transcripts that are expected to contain unstructured repeats[64]. The specific recognition of an unstructured RNA repeat requires a distinct mechanistic solution for RNA substrate discrimination. Remarkably, crystallographic studies of the Cas6 protein from *P. furiosus* have revealed the same duplicated ferredoxin-like fold observed in the Cas6e protein, but with a different mode of RNA recognition involving the opposite face of the protein (Fig. 4). In Cas6, the two ferredoxin-like folds clamp the 5′ end of the single-stranded RNA repeat sequence in place[21]. Although the RNA in this structure is disordered in the enzyme active site, biochemical studies have shown that cleavage occurs 8 nucleotides upstream of the spacer sequence[16,64]. While the nucleotide sequences at the cleavage site vary for each of the different Cas6 proteins, all Cas6 family endoribonucleases cleave their cognate RNA 8 nucleotides upstream of the spacer sequence using a metal-ion-independent mechanism.

Despite advances in our understanding of crRNA biogenesis, the diversity of *cas* genes has obscured identification of the protein factors responsible for CRISPR RNA processing in some systems. Type II immune systems consist of four *cas* genes, none of which have a detectable sequence similarity to known CRISPR-specific endoribonucleases. Recently, a different CRISPR RNA processing mechanism has been reported that involves RNase-III-mediated cleavage of double-stranded regions of the CRISPR RNA repeats[17]. The first indication of this mechanism came from deep sequencing of RNA from *S. pyogenes*. An abundant transcript containing a 25-nucleotide sequence that was complementary to the CRISPR repeat was identified. This RNA, termed tracrRNA (*trans*-activating CRISPR RNA), is coded on the opposite strand and just upstream of the CRISPR locus. Genetic and biochemical experiments demonstrated that tracrRNA and pre-crRNA are co-processed by RNase III, which produces cleavage products with a 2 nucleotide 3′ overhang[17]. *In vivo* processing of CRISPR RNAs required Cas9 (previously known as Csn1), although a precise role for this enzyme in RNA processing has not yet been defined. The essential role of cellular proteins that are not solely involved in CRISPR-mediated defence, such as RNase III, indicates that different host factors may be involved as ancillary components of these immune systems.

crRNA-guided interference

The third stage of CRISPR-mediated immunity is target interference (Fig. 2). Here crRNAs associate with Cas proteins to form large CRISPR-associated ribonucleoprotein complexes that can recognize invading

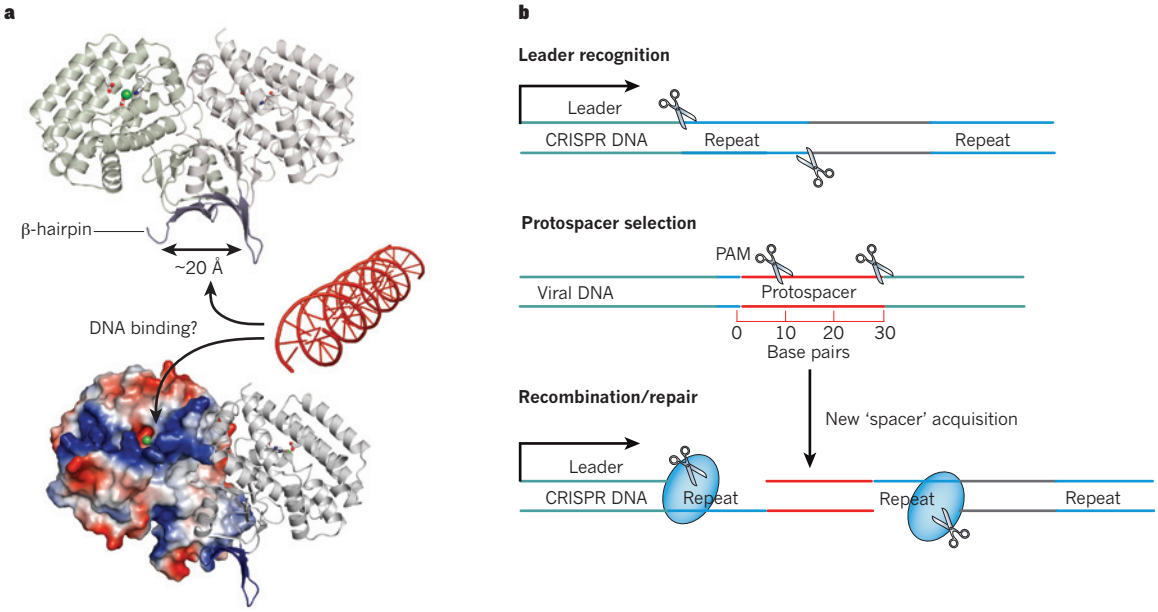

Figure 3 | Steps leading to new spacer integration. a, The Cas1 protein forms a stable homodimer where the two molecules (green and grey) are related by a pseudo-two-fold axis of symmetry (PBD ID: 3GOD)[54,56]. This organization creates a saddle-like structure in the N-terminal domain, in which β-hairpins (blue) from each symmetrically related molecule hang (like stirrups) that are separated by approximately 20 Å, and may interact with the phosphodiester backbone of double-stranded DNA. An electrostatic surface representation (bottom) reveals a cluster of basic residues (blue) that form a positively charged strip across the metal-binding surface of the C-terminal domain. This strip may serve as an electrostatic trap that positions DNA substrates proximally to catalytic metal ions (green sphere). **b**, CRISPR adaptation occurs by integrating fragments of foreign nucleic acid preferentially at the leader end of the CRISPR, forming new repeat-spacer units in the process. Protospacers are chosen non-randomly and may be selected from regions flanking the protospacer adjacent motif (PAM). Coordinated cleavage of the foreign DNA and integration of the protospacer into the leader-end of the CRISPR occurs through a mechanism that duplicates the repeat sequence and thus preserves the repeat-spacer-repeat architecture of the CRISPR locus. Although the protein components required for this process have not been conclusively identified, Cas1 and other general recombination or repair factors have been implicated (blue ovals)[32,54,56].

nucleic acids. Foreign nucleic acids are identified by base-pairing interactions between the crRNA spacer sequence and a complementary sequence from the intruder. Phage- and plasmid-challenge experiments performed in several model systems have demonstrated that crRNAs complementary to either the coding or the non-coding strand of the invading DNA can provide immunity[14,22,47,60,65,66]. This is indicative of an RNA-guided DNA-targeting system, and indeed a pathway for DNA silencing has recently been demonstrated in *S. thermophilus*[15]. DNA sequencing and Southern blots indicated that both strands of the target DNA are cleaved within the region that is complementary to the crRNA spacer sequence[15]. This mechanism efficiently eliminates foreign DNA sequences, which have been specified by the spacer region of the crRNA, but avoids targeting the complementary DNA sequences in the CRISPR region of the host chromosome. The mechanism for distinguishing self from non-self is built into the crRNA. The spacer sequence of each crRNA is flanked by a portion of the adjacent CRISPR repeat sequence, and any complementarity beyond the spacer into the adjacent repeat region signals self and prevents the destruction of the host chromosome[67].

However, not all CRISPR systems target DNA. *In vitro* experiments using enzymes from the type III-B CRISPR system of *P. furiosus* have shown that this system cleaves target RNA rather than DNA[48]. All DNA targeting systems encode a complementary DNA sequence for each crRNA in the CRISPR locus and therefore require a mechanism for distinguishing self (CRISPR locus) from non-self (invading DNA). In contrast, systems that target RNA may not be required to make this distinction because most CRISPR loci are transcribed only in one direction and thus do not generate complementary RNA targets. CRISPR systems that target RNA may be uniquely capable of defending against viruses that have RNA-based genomes. However, adaptation of the CRISPR in response to a challenge by an RNA-based virus will probably require the invading RNA to be reverse-transcribed into DNA before it can be integrated into the CRISPR locus.

Cas proteins directly participate in target binding. Recent biochemical studies have shown that CRISPR-associated complexes facilitate target recognition by enhancing sequence-specific hybridization between the CRISPR RNA and complementary target sequences[27]. A short high-affinity binding site located at one end of the crRNA spacer sequence governs the efficiency of target binding, and viruses that acquired a single mismatch in this region were able to escape detection by the immune system[60]. This high-affinity binding site is functionally analogous to the 'seed' sequence (Fig. 1) that has been identified in eukaryotic microRNAs (miRNAs)[68]. Structural and biochemical studies have shown that Argonaute proteins facilitate target recognition by pre-ordering the nucleotides at the 5′ end of the miRNA in a helical configuration[69]. This pre-ordering reduces the entropic penalty that is associated with helix formation and provides a thermodynamic advantage for target binding within this region. A similar mechanism may occur during crRNA target binding, providing an interesting example of convergent evolution between CRISPR-based immunity in prokaryotes and RNAi in eukaryotes (Fig. 1).

Structural and biochemical studies have been performed on CRISPR-associated complexes isolated from three different type I CRISPR systems[24–27,48]. These complexes seem to share some general morphological features, but the precise special arrangement of the Cas proteins and their interactions with the crRNA have been unclear. Sub-nanometre-resolution structures of the CRISPR-associated complex from *E. coli* (Cascade) have recently been determined using cryo-electron microscopy[26]. This complex is comprised of an unequal stoichiometry of 5 functionally essential Cas proteins and a 61-nucleotide crRNA[22,24,26]. The structure reveals a sea-horse-shaped architecture in which the crRNA is displayed along a helical arrangement of protein subunits that protect the crRNA from degradation[26]. The 5′ and 3′ ends of the crRNA form unique structures

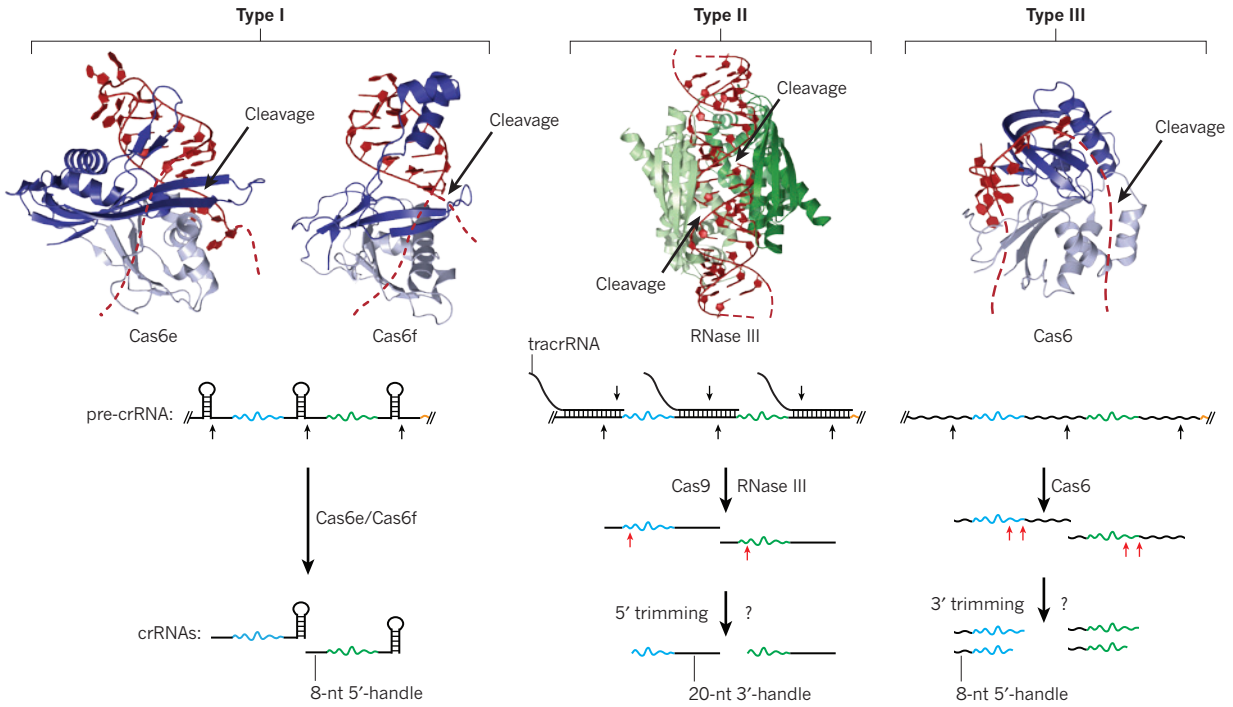

Figure 4 | Diverse mechanisms of CRISPR RNA biogenesis. CRISPR RNA repeats are specifically recognized and cleaved by diverse mechanisms. In type I CRISPR systems, Cas6e (PDB ID: 2Y8W) and Cas6f (PDB ID: 2XLK) recognize the major groove of the crRNA stem-loop primarily through electrostatic interactions using a β-hairpin and α-helix, respectively[18,19,20]. Cleavage occurs at the double-stranded–single-stranded junction (black arrows), leaving an 8-nt 5′-handle on mature crRNAs. In type II CRISPR systems, tracrRNA hybridizes to the pre-crRNA repeat to form duplex RNAs that are substrates for endonucleolytic cleavage by host RNase III (PDB ID: 2EZ6), an activity that may also require Cas9 (ref. 17). Subsequent trimming (red arrows) by an unidentified nuclease removes leftover repeat sequences from the 5′ end. Cas6 (PDB ID: 3PKM) in type III-B CRISPR systems specifically recognizes single-stranded RNA, upstream of the scissile phosphate, on a face of the protein opposite that of the previously identified active site residues[16,21,64]. The remainder of the repeat substrate probably wraps around the protein (red dashed line) to allow cleavage 8 nucleotides upstream of the repeat-spacer junction. Subsequent 3′ trimming (red arrows) generates mature crRNAs of two discrete lengths. The N-terminal domain of all Cas 6 family proteins adopts a ferredoxin-like fold (light blue).The C-terminal domain of Cas6 and Cas6e also adopts a ferredoxin-like fold but the C-terminal domain of Cas6f is structurally distinct (dark blue).

that are anchored at opposite ends of the Cascade complex, displaying the 32-nucleotide spacer sequence for base-pairing with complementary targets.

The structure of Cascade bound to a 32-nucleotide target sequence[26] reveals a concerted conformational change that could be a signal for recruiting Cas3. Cas3 — the *trans*-acting nuclease of type I CRISPR systems — may function as a target 'slicer' in a similar way to Argonaute in RNAi pathways[22,44,46,70]. Although Cas3 was implicated previously in the process of self versus non-self discrimination, recent studies have demonstrated that Cascade recognizes the PAM directly and that mutations in the PAM decrease Cascade's affinity for the target[60]. The importance of the PAM is highlighted by the recovery of phage and plasmid escape mutants, which frequently contain a single mutation in the PAM[15,51–53,60]. The structure of Cascade indicates that the PAM is positioned near the 'tail' of the sea-horse-shaped complex. High-resolution structures and mutational analysis of the nucleic acid and protein components in this and related systems are needed to determine the mechanisms of target authentication and degradation.

Applications of CRISPR structure and function

The sequence diversity of CRISPR loci, even within closely related strains, has been used for high-resolution genotyping and forensic medicine. This technique, known as spoligotyping (spacer oligotyping), has been applied successfully to the analysis of human pathogens, including *Mycobacterium tuberculosis*[71], *Corynebacterium diphtheriae*[72] and *Salmonella enterica*[73]. Spoligotyping was developed long before the function of CRISPRs was understood, but now that studies have begun to reveal the biological function and mechanism of CRISPR-mediated genetic silencing, new opportunities for creative applications have emerged. Laboratory strains of bacteria are grown in high-density bioreactors for many different applications in the food industry, and they are becoming increasingly important in the production of biofuels. CRISPR systems offer a natural mechanism for adapting economically important bacteria for resistance against multiple phages.

The biochemical activities of various Cas proteins may have useful applications in molecular biology in much the same way that DNA restriction enzymes have revolutionized cloning and DNA manipulation. A wide range of CRISPR-specific endoribonucleases that recognize small RNA motifs with high affinity expand the number of tools available for manipulating nucleic acids. In addition, a crRNA-guided ribonucleoprotein complex in *P. furiosus* was shown to cleave target RNAs[48]. Site-specific cleavage of target RNA molecules could have a range of uses, from generating homogeneous termini after *in vitro* transcription to targeting a specific intracellular messenger RNA for inactivation in a similar way to RNAi. CRISPRs also provide a new mechanism for limiting the spread of antibiotic resistance or the transfer of virulence factors by blocking horizontal gene transfer[15,47]. In addition, CRISPRs participate in a regulatory mechanism that alters biofilm formation in *P. aeruginosa*[74,75]. Although the clinical relevance of CRISPRs remains to be demonstrated, the opportunities for creative implementation of this new gene-regulation system are perceivably vast.

Future directions of CRISPR biology

The discovery of some of the fundamental mechanisms of CRISPR-based adaptive immunity has raised new questions and highlighted the areas with the greatest potential for future research. Although CRISPR RNA processing and targeting steps are now understood in some detail, how and when target sequences are identified during a phage infection

or plasmid transformation are still unclear. Furthermore, why DNA or RNA target sequences are chosen, and their fate once they are bound to a crRNA-targeting complex is not well understood. In addition, the mechanisms by which foreign sequences are selected and integrated into CRISPR loci are almost entirely unknown. Some CRISPR loci seem to be considerably more active than others, at least under laboratory conditions, so selection of the model organisms will be important. The diversity and prevalence of CRISPR systems throughout microbial communities ensures that new findings and applications in this field will be forthcoming in the years ahead. ■

1. Hoskisson, P. A. & Smith, M. C. Hypervariation and phase variation in the bacteriophage 'resistome'. *Curr. Opin. Microbiol*. **10,** 396–400 (2007).
2. Rodriguez-Valera, F. *et al*. Explaining microbial population genomics through phage predation. *Nature Rev. Microbiol.* **7,** 828–836 (2009).
3. Weinbauer, M. G. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* **28,** 127–181 (2004).
4. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature Rev. Microbiol.* **8,** 317–327 (2010).
5. Stern, A. & Sorek, R. The phage-host arms race: shaping the evolution of microbes. *Bioessays 33,* 43–51 (2010).
6. Al-Attar, S., Westra, E. R., van der Oost, J. & Brouns, S. J. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. *Biol. Chem.* **392,** 277–289 (2011).
7. Deveau, H., Garneau, J. E. & Moineau, S. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.* **64,** 475–493 (2010).
8. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327,** 167–170 (2010).
9. Karginov, F. V. & Hannon, G. J. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell* **37,** 7–19 (2010).
10. Makarova, K. S. et al. Evolution and classification of the CRISPR-Cas systems. *Nature Rev. Microbiol.* **9,** 467–477 (2011).
11. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Rev. Genet.* **11,** 181–190 (2010).
12. Sorek, R., Kunin, V. & Hugenholtz, P. CRISPR-a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Rev. Microbiol.* **6,** 181–186 (2008).
13. Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320,** 1047–1050 (2008).
14. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).
This study provided the first direct evidence for CRISPR immune system function, demonstrating that new spacer acquisition provides resistance against future phage infection in a *cas* gene-dependent manner.
15. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).
This study showed that CRISPRs can acquire new spacers upon plasmid challenge and provided the first example of crRNA-guided cleavage of double-stranded DNA *in vivo*.
16. Carte, J., Wang, R., Li, H., Terns, R. M. & Terns, M. P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22,** 3489–3496 (2008).
17. Deltcheva, E. et al. CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (2011).
This study identified a new CRISPR RNA processing pathway that involves a *trans*-acting RNA complementary to the CRISPR repeat sequence and demonstrated that processing of this duplex is mediated by cellular RNase III.
18. Gesner, E. M., Schellenberg, M. J., Garside, E. L., George, M. M. & MacMillan, A. M. Recognition and maturation of effector RNAs in a CRISPR interference pathway *Nature Struct. Mol. Biol.* **18,** 688–692 (2011).
19. Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J. A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329,** 1355–1358 (2010).
20. Sashital, D. G., Jinek, M. & Doudna, J. A. An RNA induced conformational change required for CRISPR RNA cleavage by the endonuclease Cse3. *Nature Struct. Mol. Biol.* **18,** 680–687 (2011).
21. Wang, R., Preamplume, G., Terns, M. P., Terns, R. M. & Li, H. Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure* **19,** 257–264 (2011).
22. Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321,** 960–964 (2008).
This paper revealed that long CRISPR RNA transcripts are processed into small RNAs (crRNAs) by a dedicated endoribonuclease and that the processed RNAs are bound by a Cas protein complex (Cascade), which together with Cas3 are required for phage protection.
23. Hale, C., Kleppe, K., Terns, R. M. & Terns, M. P. Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *RNA* **14,** 2572–2579 (2008).
24. Jore, M .M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Struct. Mol. Biol.* **18,** 529–536 (2011).
25. Lintner, N. G. *et al.* Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (crispr)-associated complex for antiviral defense (CASCADE). *J. Biol. Chem.* **286,** 21643–21656 (2011).
26. Wiedenheft, B. et al. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477,** 486–489 (2011).
This paper presented the first sub-nanometer resolution structures of Cascade, showing how the crRNA is accommodated within a large ribonucleoprotein complex that is involved in foreign nucleic acid surveillance.
27. Wiedenheft, B. et al. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl Acad. Sci. USA* **108,** 10092–10097 (2011).
28. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169,** 5429–5433 (1987).
29. Bolotin, A., Ouinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–2561 (2005).
30. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60,** 174–182 (2005).
31. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151,** 653–663 (2005).
32. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. 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 (2006).
33. Grissa, I., Vergnaud, G. & Pourcel, C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **35,** W52–W57 (2007).
34. Rousseau, C., Gonnet, M., Le Romancer, M. & Nicolas, J. CRISPI: a CRISPR interactive database. *Bioinformatics* **25,** 3317–3318 (2009).
35. Kunin, V., Sorek, R. & Hugenholtz, P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* **8,** R61 (2007).
36. Bland, C. et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8,** 209 (2007).
37. Dsouza, M., Larsen, N. & Overbeek, R. Searching for patterns in genomic data. *Trends Genet.* **13,** 497–498 (1997).
38. Edgar, R. C. PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* **8,** 18 (2007).
39. Jansen, R., Embden, J. D., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43,** 1565–1575 (2002).
40. Pougach, K. et al. Transcription, processing and function of CRISPR cassettes in Escherichia coli. *Mol. Microbiol.* **77,** 1367–1379 (2010).
41. Pul, U. *et al.* Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Mol. Microbiol.* **75,** 1495–1512 (2010).
42. Westra, E. R. *et al.* H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.* **77,** 1380–1393 (2010).
43. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **1,** e60 (2005).
44. Sinkunas, T. *et al.* Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* **30,** 1335–1342 (2011).
45. Han, D. & Krauss, G. Characterization of the endonuclease SSO2001 from Sulfolobus solfataricus P2. *FEBS Lett.* **583,** 771–776 (2009).
46. Mulepati, S. & Bailey, S. 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,** 31896–31903 (2011).
47. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in *staphylococci* by targeting DNA. *Science* **322,** 1843–1845 (2008).
48. Hale, C. R. *et al.* RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139,** 945–956 (2009).
Most CRISPR systems appear to target DNA, but this article reported the isolation of a multisubunit ribonucleoprotein complex (Cmr-complex) from *P. furiosus* that cleaves target RNAs (not DNA) at a fixed distance from the 3′ end of the crRNA-guide.
49. Tyson, G. W. & Banfield, J. F. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* **10,** 200–207 (2008).
50. Snyder, J. C., Bateson, M. M., Lavin, M. & Young, M. J. Use of cellular CRISPR (clusters of regularly interspaced short palindromic repeats) spacer-based microarrays for detection of viruses in environmental samples. *Appl Environ Microbiol* **76,** 7251–7258 (2010).
51. Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190,** 1390–1400 (2008).
52. Horvath, P. *et al.* Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* **190,** 1401–1412 (2008).
53. Sapranauskas, R. *et al.* The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **39,** 9275–9282 (2011).
54. Babu, M. *et al.* A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Mol. Microbiol.* **79,** 484–502 (2011).
55. Han, D., Lehmann, K. & Krauss, G. SSO1450--a CAS1 protein from *Sulfolobus*

solfataricus P2 with high affinity for RNA and DNA. *FEBS Lett.* **583,** 1928–1932 (2009).
56. Wiedenheft, B. *et al.* Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated antiviral defense. *Structure* **17,** 904–912 (2009).
57. Chen, C. S. *et al.* A proteome chip approach reveals new DNA damage recognition activities in *Escherichia coli*. *Nature Methods* **5,** 69–74 (2008).
58. Aguilera, A. The connection between transcription and genomic instability. *EMBO J.* **21,** 195–201 (2002).
59. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155,** 733–740 (2009).
60. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl Acad. Sci. USA* **108,** 10098–10103 (2011).
61. Tang, T. H. *et al.* Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl Acad. Sci. USA* **99,** 7536–7541 (2002).
62. Tang, T. H. *et al.* Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* **55,** 469–481 (2005).
63. Ebihara, A. *et al*. 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,** 1494–1499 (2006).
64. Carte, J., Pfister, N. T., Compton, M. M., Terns, R. M. & Terns, M. P. Binding and cleavage of CRISPR RNA by Cas6. RNA **16,** 2181–2188 (2010).
65. Gudbergsdottir, S. *et al.* Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* **79,** 35–49 (2011)
66. Manica, A., Zebec, Z., Teichmann, D. & Schleper, C. *In vivo* activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. *Mol. Microbiol.* **80,** 481–491 (2011).
67. Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463,** 568–571 (2010).
This study identified a mechanism for distingishing self from non-self, which relies on base-pairing potential in regions outside the crRNA spacer sequence.
68. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116,** 281–297 (2004).
69. Parker, J. S., Parizotto, E. A., Wang, M., Roe, S. M. & Barford, D. Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol. Cell* **33,** 204–214 (2009).
70. Beloglazova, N. *et al.* Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO J.* **30,** 4616–4627 (2011).
71. Groenen, P. M., Bunschoten, A. E., van Soolingen, D. & van Embden, J. D. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol. Microbiol.* **10,** 1057–1065 (1993).
72. Mokrousov, I., Limeschenko, E., Vyazovaya, A. & Narvskaya, O. *Corynebacterium diphtheriae* spoligotyping based on combined use of two CRISPR loci. *Biotechnol. J.* **2,** 901–906 (2007).
73. Liu, F. *et al.* Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. *Appl. Environ. Microbiol.* **77,** 1946–1956 (2011).
74. Cady, K. C. & O'Toole, G. A. Non-identity targeting of yersinia-subtype CRISPR-prophage interaction requires the Csy and Cas3 proteins. *J. Bacteriol.* **193,** 3433–3445 (2011).
75. Zegans, M. E. *et al.* Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of Pseudomonas aeruginosa. *J. Bacteriol.* **191,** 210–219 (2009).
76. Obbard, D. J., Gordon, K. H. J., Buck, A. H. & Jiggins, F. M. The evolution of RNAi as a defence against viruses and transposable elements. *Philos. Trans. R. Soc. B Biol. Sci.* **364,** 99–115 (2009).
77. Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318,** 761–764 (2007).
78. Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D-melanogaster germline. *Curr Biol* **11,** 1017–1027 (2001).
79. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136,** 215–233 (2009).
80. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466,** 835–840 (2010).

Acknowledgements B.W. is a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation. S.H.S. acknowledges support from the National Science Foundation and National Defense Science & Engineering Graduate Research Fellowship programs. J.A.D. is an Investigator of the Howard Hughes Medical Institute.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to J.A.D. (doudna@berkeley.edu).